University of Dundee

DOCTOR OF PHILOSOPHY

The Use of Texture Analysis in Breast Magnetic Resonance Imaging

Waugh, Shelley Ann

Award date:
2014

The Use of Texture Analysis in Breast Magnetic Resonance Imaging

Shelley Ann Waugh

2014

University of Dundee

THE USE OF TEXTURE ANALYSIS

IN

BREAST MAGNETIC RESONANCE IMAGING

SHELLEY ANN WAUGH

A THESIS PRESENTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

UNIVERSITY OF DUNDEE

MAY 2014

Chapter One- Introduction

Chapter Two- Theory

Chapter Three- Phantom Validation Study

Chapter Four- Retrospective Patient Study

Chapter Five- Comparison of Texture Analysis Techniques

Chapter Six-Prospective Patient Study

Chapter Seven- Texture Analysis in Treatment Response

Chapter Eight- Conclusions and Future Work

Appendix A- Mathematical description of texture features

Appendix B- Program coding used for max-min analysis

Appendix C- Caldicott Approval

Appendix D- Publications and presentations

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

This work has at times been a lifeline, it's been a labour of love and has been completed, partly through stubbornness, but mainly due to the unending support from countless people who wouldn't let me give up when times were tough, and never stopped believing I would achieve my goal.

To my colleagues, family and friends who supported and encouraged me, and kept me going when I felt I'd had enough, thank you. To those who continually asked 'How is the PhD?' I'm sorry for the times I groaned 'Don't ask!'. To those who took me out to escape when I needed a break, it's because of you I managed to complete this!

To my two supervisors- Professor Lerski and Professor Thompson… there aren't enough words to express my gratitude for your guidance, your support and your belief in me when things weren't going quite as I'd planned. I couldn't have hoped for better supervisors and thank you for the faith you put in me that I could make it, and that my work was to be proud of.

Finally, and most importantly, thank you to my mum and dad. You've supported me from day one and pushed me to achieve my maximum. You've given me pep talks, stern words and encouragement when life got on top of me and never once doubted I would get here. I promise this is the end of studying…for now!

Mum, dad; this is for you both.

DECLARATION

I, Shelley Waugh, declare that I am the sole author of this thesis and the results presented herein are from research I have carried out and have not previously been accepted for a higher degree. All references cited within this work have been consulted and work pertaining to these references has been appropriately acknowledged.

Signed: SWaugh

Date: 12th May 2014.

SUPERVISOR STATEMENT

I, Professor Richard Lerski, confirm that Shelley Waugh has spent the equivalent of at least 9 terms within the Access Division Directorate and that she has fulfilled the conditions of the University of Dundee therby qualifying her to submit her thesis in the application for the degree of Doctor of Philosophy

Signed: Richard A Lerski

Date: 10/2/14

SYMBOLS AND ABBREVIATIONS

α	Flip angle
α_e	Ernst angle
γ	Gyromagnetic ratio
κ	Inter-observer agreement statistic
μ	Grey level mean
θ	Interpixel direction (COM model)
θ_A	Parameter weights in ARM model
σ_A	Noise parameter in ARM model
σ	Grey level standard deviation
ω_L	Larmor frequency
ANOVA	Analysis of variance (statistic)
ARM	Auto-regressive model
B_0	Static magnetic field
B_1	RF excitation field
BI-RADS	Breast imaging reporting and data system
BRCA1/2	Gene mutation resulting in increased breast cancer risk
BW	Bandwidth
CAD	Computer aided detection
COM	Co-occurrence matrix
CoR	Coefficient of repeatability
CoV	Coefficient of variation
CR	Complete response
CT	Computed tomography
d	Inter-pixel distance (COM model)
DBT	Digital breast tomosynthesis
DCIS	Ductal carcinoma *in situ*
DICOM	Digital Imaging and communications in medicine
DOC	Docetaxyl (chemotherapy agent)
Ent	Entropy
ER	Oestrogen receptor
FEC	Fluoracil, epirubicin, cyclophosphamide (chemotherapy regime)
FFT	Fast fourier transform
FID	Free induction decay
FLASH	Fast low angle shot
FOV	Field of view
Gd2/3	Grade 2/3
GE	Gradient echo
GRA	Gradient features
HER2	Human epidermal growth factor 2 receptor
Hz	Hertz
IHC	Immunohistochemistry
IR	Inversion recovery
k-NN	k- nearest neighbour
LL/LH/HL/HH	filters in WAV model (L- low pass, H- high pass)

M	Bulk magnetisation
M_0/M_z	Magnetisation in direction of static field
M_{xy}	Magnetisation in transverse plane
MHz	Mega hertz
MMWP	Multi-modality workplace (Siemens workstation)
MRD	Minimal residual disease
MRI	Magnetic resonance imaging
ms	Milliseconds
NAC	Neoadjuvant chemotherapy
NHS	National Health Service
NMR	Nuclear magnetic resonance
NPV	Negative predictive value
pCR	Pathological complete response
PD	Proton density
PD (RECIST)	Progressive disease
PPV	Positive predictive value
PR	Progesterone receptor
PR (RECIST)	Partial response
RCB	Residual cancer burden
RCB-I	Minimal residual disease
RCB-II	Moderate residual disease
RCB-III	Extensive residual disease
RECIST	Response evaluation criteria in solid tumours
RF	Radiofrequency
RLM	Run length matrix
ROC	Receiver operating characteristic
ROI	Region of interest
SD	Stable disease
SE	Spin echo
sigfigs	Significant figures
SNR	Signal to noise
T	Tesla
T_1	Spin-lattice / longitudinal relaxation time
T_2	Spin-spin / transverse relaxation time
TA	Texture analysis
T_c	Critical temperature
TDM-1	Trastuzumab emtansine (chemotherapy agent)
TE	Echo time
TNBC	Triple negative breast cancer
TR	Repetition time
TRA	Trastuzumab (chemotherapy agent)
TSE	Turbo spin echo
US	Ultrasound
WAV	Wavelet features
Δz	slice thickness

SUMMARY

Breast cancer is the most common cancer in the UK, accounting for a third of cancers diagnosed annually. To better manage patients it is essential to diagnose early, with high accuracy, the type and grade of cancer to ensure correct and rapid treatment.

Breast magnetic resonance imaging (MRI) provides excellent soft tissue contrast and is unaffected by fibroglandular tissue density. By dynamically imaging during the injection of a contrast agent, a sensitivity approaching 100% is realised and lesion vascularity is portrayed. This makes the examination not only useful in detection and staging of breast cancer, but also for monitoring response to neoadjuvant chemotherapy (NAC) treatment.

This work used an image processing technique known as texture analysis (TA), which analyses pixel intensity distributions on a pixel-by-pixel scale to identify patterns that may not be visually interpretable, to probe MRI images of women with proven breast cancer. Whilst previous works have demonstrated differentiation between normal, benign and malignant tissue, this work sought to extend this and look at classification of breast cancer subtypes, the utility in a clinical environment and to assess whether the technique could identify early response in patients undergoing NAC.

TA cannot only be applied using standard MRI set-ups, but the studies demonstrated preliminary promise in the classification of different cancer subtypes- both in terms of histological subtype and grading, as well as state-of-the-art molecular subtyping. While larger patient data sets are required to demonstrate this definitively, initial results show encouraging findings.

It has also been shown that TA can be used in patients undergoing NAC to indicate whether the patient will respond well or not, and of particular interest is that these results appear to correlate well with the final pathological outcome.

The research within this thesis has clearly demonstrated that TA is a useful research tool within the area of breast MRI and further investigation in this area is essential.

CHAPTER ONE: INTRODUCTION

1.1 THE HEALTHY BREAST

1.1.1 Anatomy of the Breast

The function of the female breast is for milk production in order to suckle and nourish the young. Figure 1.1 shows a schematic diagram of the structure of the breast [1]. The breast develops during puberty, under the influence of hormones such as oestrogen and progesterone [2]. The breast tissue (often referred to as fibroglandular tissue or parenchyma) consists of 15-20 glandular lobes which form a conical mass with the nipple at the apex, and each lobe has a lactiferous duct through which milk drains and collects in lactiferous sinuses, just below the surface of the nipple. The lactiferous ducts are formed from

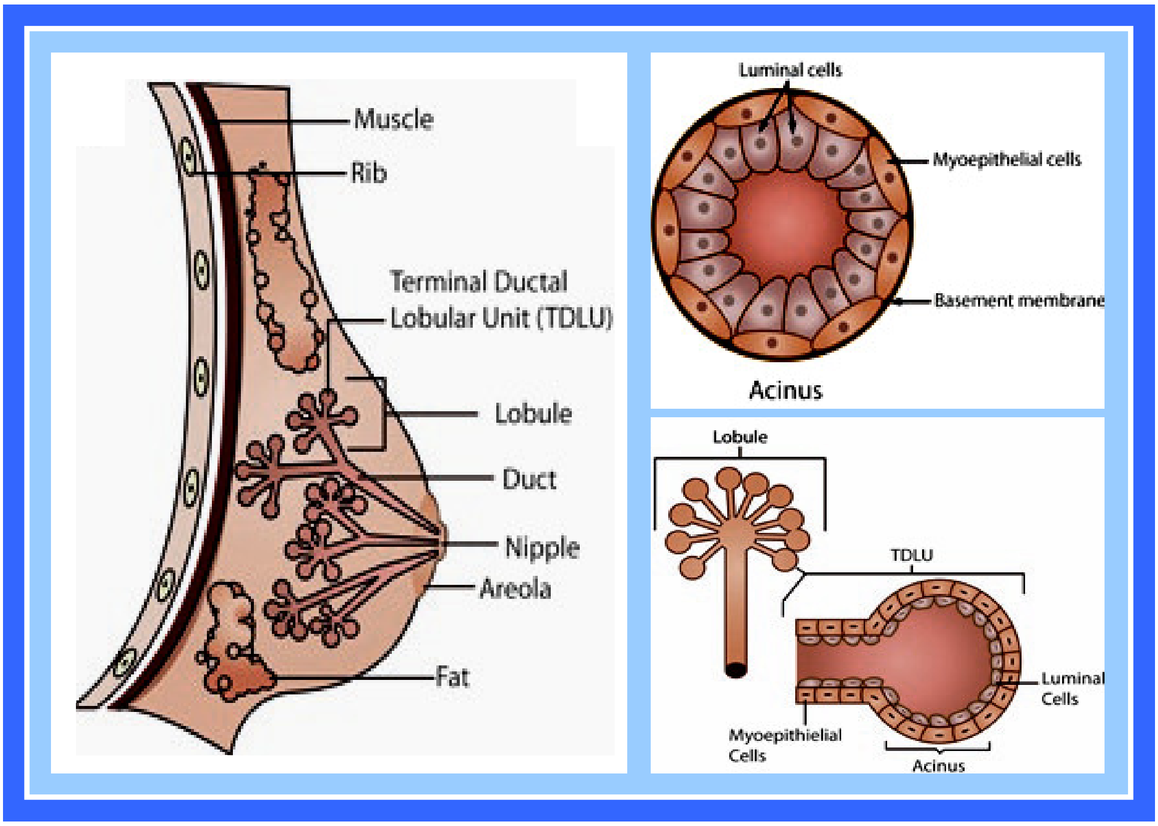


Figure 1.1- Schematic labelled view of the healthy breast [1]

epithelial cells, Figure 1.1. Each lobe is formed from 20-40 lobules [3] consisting of branching ducts leading back to the milk producing glands known as alveoli, or acini. The glandular tissue and ducts are supported by connective tissue (Cooper's ligaments) which extend over the pectoralis muscle and the skin, and are surrounded by fat which gives the breast its form.

1.1.2 Normal Changes

In pre-menopausal women, the breast is responsive to cyclical changes in oestrogen and progesterone as controlled by the menstrual cycle as shown in Figure 1.2 [4]. During the second half of the cycle, the high levels of oestrogen and progesterone result in increased proliferative activity within the luminal epithelial cells [5-7]. This, coupled with vascular and lymphatic changes results in an increased breast volume of up to 100 ml [8] which is attributed to not only the increase in parenchymal volume, but an accompanying increase in water

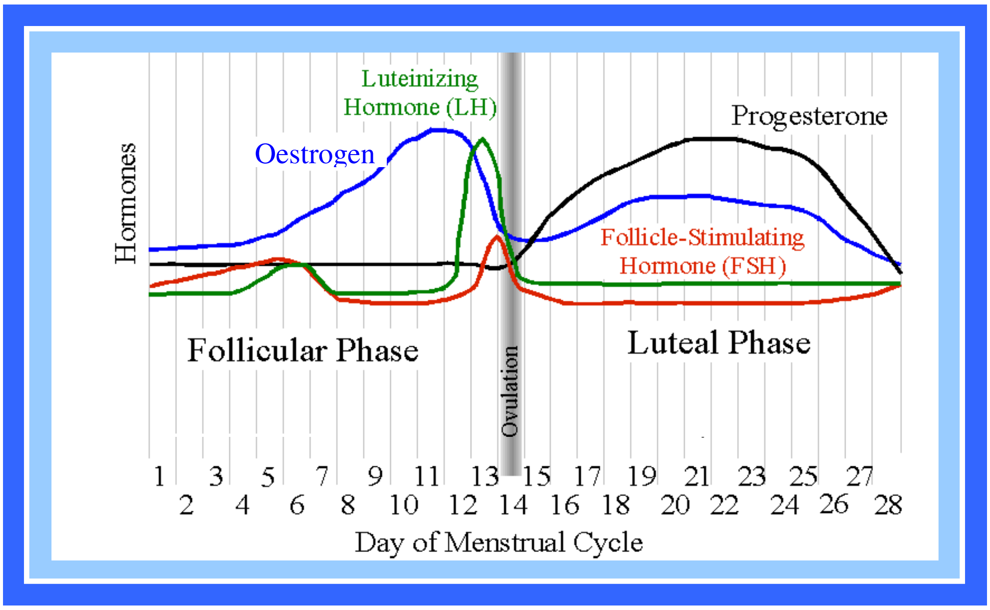


Figure 1.2- Hormonal fluctuations throughout the menstrual cycle

content [9] due to stromal oedema [5]. At the end of the cycle, apoptosis occurs whereby the epithelium atrophies and the lumina of the alveoli close [5, 6] resulting in ducts with little or no lumen [6].

In post menopausal women, there is no longer the cyclical influence of oestrogen and progesterone on breast tissue and therefore the glandular tissue shrinks and is replaced by fatty tissue. The epithelial cells atrophy and the connective ligaments generally lose strength and stretch.

1.1.3 Benign Breast Disease

Benign breast disease is that which poses no threat to life, despite often presenting with symptoms that may be concerning, such as lumpiness, pain or mass. Benign disease is split into two categories- that which poses no increased risk of breast cancer development and that which can result in an increased risk of subsequent breast cancer development.

The most common type of benign change with no increased threat of breast cancer development is benign ‘fibrocystic disease’. This change shows no increased proliferation of epithelial cells and is believed to be caused by a hormone imbalance and the most common findings are mammary cysts, stromal fibrosis and lobule expansion [1].

Some benign conditions that can result in an increased risk of developing breast cancer are atypical ductal hyperplasia, radial scar and papilloma.

1.1.4 *In Situ* Carcinoma

Non-invasive cancers do not invade into normal surrounding breast tissues and are characterised by hyperplasia either in ducts (Ductal Carcinoma *in situ*, DCIS) or lobules (Lobular Carcinoma *in situ*, LCIS) with a disordered proliferative appearance [10]. DCIS increases the risk of developing breast cancer later in life and is termed pre-cancerous disease, whereas LCIS is a misleading term as there is no direct link with pre-malignancy.

1.2 BREAST CANCER

Breast cancer is the most common cancer in the United Kingdom with more than 120 new women diagnosed per day. Cases have increased by 14% over the past ten years and it is estimated that by 2024 the number of women presenting annually with breast cancer will have reached 55, 700 [11].

Mortality has fallen over the past 30 years, partly due to improved treatment options as well as the introduction of the National Health Service (NHS) Breast Screening programme, which often detects early onset of malignancy before tumours are able to be detected by physical examination. As survival is closely related to the stage of diagnosis, earlier detection is crucial in managing the treatment of the disease.

1.2.1 Pathophysiology of Breast Cancer

While it is still unclear exactly what initiates the onset of breast cancer, certain risk factors have been identified that result in increased risk of breast cancer development, such as dense parenchymal breast tissue, early menarche, late menopause, age of first full-term pregnancy, exposure to ionising radiation, obesity, use of hormone replacement therapy, alcohol consumption, smoking and certain benign conditions as discussed in Section 1.1.3. There has also been a link with family history of breast cancer and the inheritance of the BRCA1 or BRCA2 gene mutation increases a woman's lifetime risk of breast cancer to between 40 and 85% [12]. Both of these hereditary genes result in a high prevalence of breast cancer in women of a younger age.

The progression from healthy breast to invasive, metastatic breast cancer is shown in Figure 1.3 as modified from [13]. Early stage breast cancer begins with the hyperplasia of epithelial (or luminal) cells, or the basal/myoepithelial layer in either the lactiferous duct or milk-producing lobule which has been hypothesised to be due to a hormonal imbalance over-stimulating these cells. Whilst the basal cell layer (as shown in Figure 1.1 and Figure 1.3) remains intact, this disease is classified as *in situ* cancer (either ductal carcinoma *in situ* [DCIS] or lobular carcinoma *in situ* [LCIS]) as there is no ability for invasive spread through lymphatics or the blood vessels. However both *in situ* cancers can spread along the ductal system. Not all in situ carcinoma will progress to invasive cancer- particularly low grade DCIS. It has been estimated, however, that around 40% of DCIS cases will progress to invasive breast cancer within 30-years [14].

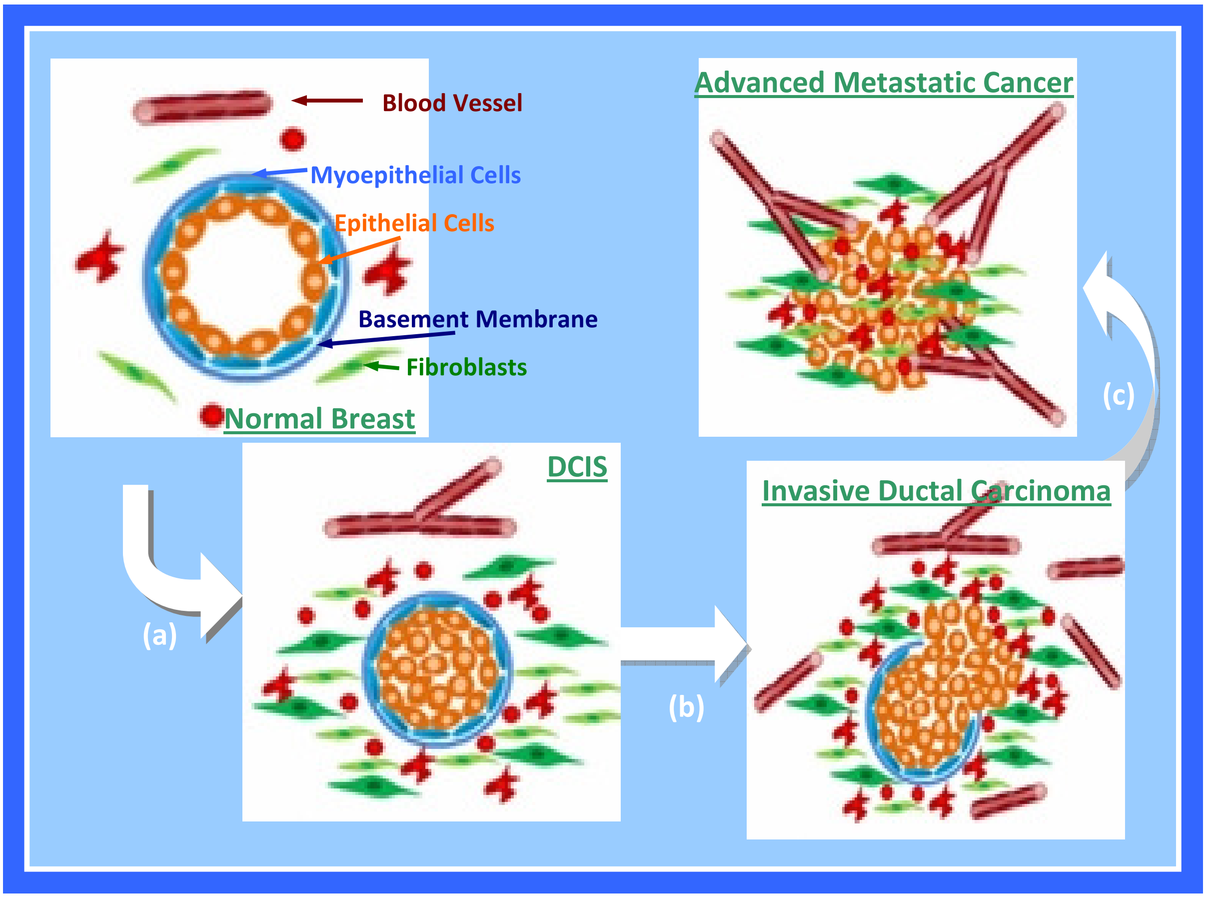


Figure 1.3- Progression of breast cancer through from the normal healthy breast to invasive, metastatic breast cancer
(a) progression to in situ cancer (b) progression to invasive cancer (c) progression to metastatic cancer

Once the basal layer is disrupted, the tumour cells are able to grow more freely, outwith the confines of the duct or lobule and the disease is then classified as invasive breast cancer (Figure 1.3(b)). The subsequent reactive proliferation of connective tissue results in dense layers of fibroblasts around the tumour and a hardening of the breast tissue- the ‘lumps’ that are usually felt on palpation. The tumour cells are free to proliferate rapidly and release an angiogenic stimulator, known as vascular endothelial growth factor (VEGF), that disrupts the basement membrane of blood vessels, causing them to sprout capilliaries that infiltrate the tumour to supply oxygen and nutrients to fuel free growth [15]. Once the tumour has acquired its own vascular system, it becomes capable of unrestricted growth and has the potential for metastasising through the lymph glands or the blood stream [16].

1.2.2 Histological Classifications of Breast Cancer

More than 95% of breast tumours arise in the milk-producing glands or ducts [17]. Breast cancer is a very heterogeneous disease and there are a wide range of histopathological types of breast cancer defined by the World Health Organization [18]. Each type of tumour has a characteristic growth pattern, and can be classified based on the appearance of cells under a microscope. This thesis will only deal with the most common histopathological types of breast cancer- invasive ductal, invasive lobular and ductal carcinoma *in situ* (DCIS), which are briefly discussed and a typical histology pattern for each shown in Figure 1.4.

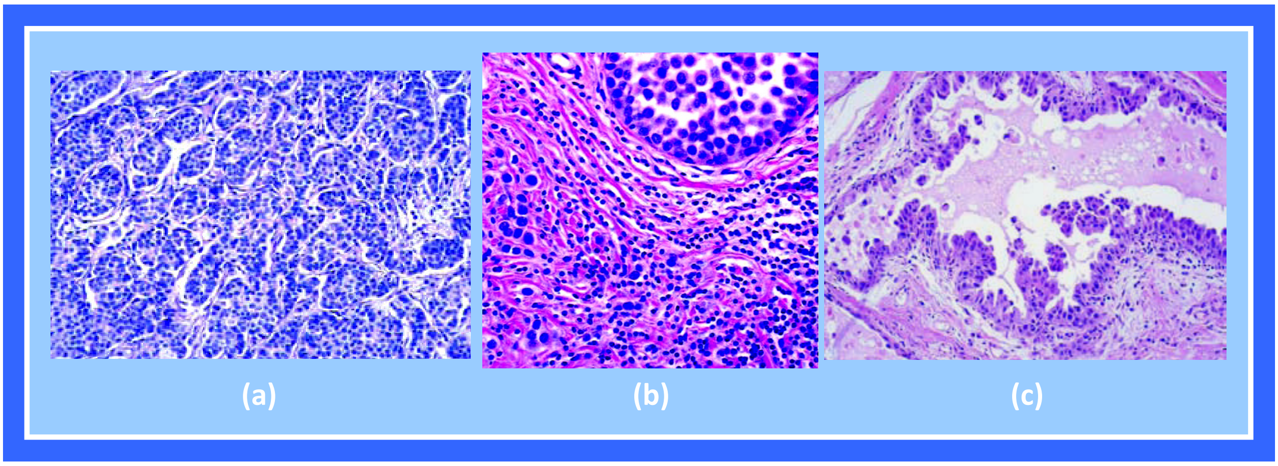


Figure 1.4- Histology for (a) Grade 2 ductal cancer (b) Grade 3 lobular cancer (c) Intermediate DCIS

Ductal cancer is the most common type of breast cancer, accounting for an estimated 80% of invasive breast cancers diagnosed in the US [19]. Ductal cancer cells typically invade stroma forming variable patterns or grow as solid sheets or nests [18].

Lobular cancers which account for around 10% of invasive cancers [19] are generally composed from small, non-cohesive cells which are dispersed through connective tissue or arranged in single-file linear patterns [18]. The majority of lobular cancers demonstrate a complete loss of the adhesion molecular e-cadherin which is attributed to the histological diffuse growth pattern [20].

DCIS, while not an invasive disease, is associated with an increased risk of subsequently developing invasive cancer. This manifests as increased epithelial proliferation within the terminal duct lobular unit. It remains bounded by the basement membrane and grading is associated with the degree of cellular atypia.

1.2.3 Grading of Breast Cancer

The grading of a cancer refers specifically to the characteristics of the cancerous cells. Grading of a lesion is standardised according to the World Health Organization classification as described in Table 1.1.

	Grading *1*	 *2*	 *3*
Degree of tubule/ glandular structure formation	Majority of tumour (>75%)	Moderate degree (10-75%)	Little or none (<10%)
Nuclear pleomorphism	Small, regular uniform cells	Increase in size and variability	Marked variation
Mitotic count	Low	Moderate	High

Table 1.1- World Health Organization classification for breast cancer grading [18]

In general Grade 1 cancers are low grade and tend to grow more slowly while Grade 3 tumours are very different from normal cells and grow much more quickly.

1.2.4 Molecular Subtypes of Breast Cancer

Recently there has been an increasing drive towards patient targeted treatments as it is recognised that the receptor status of a breast cancer can result in more successful treatment outcomes for specific cancer types. The most commonly considered receptor statuses are oestrogen (ER), progesterone (PR) and human epidermal growth factor 2 (HER2) receptors and this leads to a new classification regime. Therefore breast cancer is now more commonly becoming described by the molecular subtype [21] in terms of luminal cancers (ER positive, HER2 negative), HER2 cancers (HER2 positive) and triple negative cancers (ER negative, HER2 negative), which each exhibit distinct growth patterns [22, 23] and require different therapeutic regimens.

1.2.5 Staging of Cancer

Staging of a cancer refers to how advanced a cancer is when detected, and is often used for planning radical treatment or surgery. Breast cancer is staged using a four point scale (Table 1.2) according to how advanced the tumour is, and how far it has spread.

Stage		*Tumour*	*Physiology*		*Spread*
			Axillary Lymph Nodes?	*Clumping of nodes?*[*]	
I		<2cm	✗	✗	✗
II	*A*	<2cm	✓	✗	✗
		<5cm	✗	✗	✗
	B	<5cm	✓	✗	✗
		>5cm	✗	✗	✗
III	*A*	<5cm	✓	✓	✗
		>5cm	✓	✗	✗
	B	Fixed to skin or chest wall	✓	✓	✗
	C	Any size	✓	✓	To either the breast bone or collarbone lymph nodes
IV		Any size	✓	✓	Other parts of body

Table 1.2- Summary of the four different stages of breast cancer, according to how advanced the tumour is ([*]clumping of axillary nodes or sticking to other structures)

The earlier that breast cancer is detected, the better the prognosis for disease free survival. It has been reported that the 5-year survival rate decreases from 92% to 13% for Stage IV cancer compared to Stage I disease [11] and therefore imaging plays a crucial role in increasing survival times by detecting cancers at an earlier stage.

1.3 THE 'TRIPLE ASSESSMENT'- DIAGNOSTIC TECHNIQUES

Whilst breast cancer is the most common form of cancer in women, it is often treatable which requires identification of malignancy as well as accurate grading, staging and classification. In general, imaging, followed by a targeted biopsy, is the most reliable method of performing this task.

The usefulness of a diagnostic test is described in terms of its sensitivity, specificity and predictive values. Sensitivity refers to how well a test identifies disease in patients who truly have a disease, while specificity is concerned with how well the test identifies those without disease [24]. Predictive values are concerned with the probability of the test giving a correct diagnosis- i.e. positive predictive value (PPV) is the proportion of patients who have a positive examination who truly who have the disease, while negative predictive value (NPV) is the proportion of patients with a negative test who do not have the disease [25].

An ideal imaging examination therefore would have a high sensitivity, specificity, positive and negative predictive value such that every patient with disease was correctly diagnosed, and every patient without disease was correctly identified.

Women with suspected breast cancer will usually undergo 'triple assessment' whereby a clinical history and physical examination, ultrasound and mammography examination, subsequently followed by biopsy, will be performed. The reported diagnostic characteristics for each of these procedures are reported in Table 1.3, and each is discussed in further detail below.

	Sensitivity	*Specificity*	*PPV*
Mammography	67.8 - 77.6 %	75.0 – 98.8 %	35.8 – 85.7 %
Ultrasound	75.3 – 83.0 %	34.0 – 96.8 %	20.5 – 73.5 %
Physical Examination	27.6 – 50.3 %	92.0 – 99.4 %	28.9 – 94.0%

Table 1.3- Comparison of sensitivity, specificity and positive predictive value (PPV) for each breast examination as reported in the literature [26, 27]

1.3.1 Physical Examination

Based on a clinical history, including questions about the risk factors outlined previously, clinicians feel for signs of malignancy- typically cancer is a palpable breast mass with indistinct borders that may be attached to skin or deep fascia [28-30]. This is very much dependent on the experience of the clinician and often cysts cannot be conclusively distinguished from solid masses [28, 29]. The sensitivity of physical examination is dependent on the size of the lump and no mortality benefit has been demonstrated for this technique alone [26, 30].

1.3.2 Mammography

Mammography is one of the most common methods of imaging the breasts and is currently the primary imaging modality utilised in the NHS breast screening program. It uses low doses of x-ray radiation and breast compression to produce an image that shows the x-ray attenuation properties of the breast. Areas of attenuation different to that of normal tissue, such as architectural distortion, masses and abnormal calcification, may be indicative of malignancy (see Figure 1.5). Whilst mammography was traditionally carried out using x-ray film; recent trends have been towards digital full-field mammography (FFDM) which utilises a digital detector allowing manipulation of brightness and contrast on a computer monitor. A study comparing the two techniques in just under 43,000 women demonstrated that diagnostic accuracy was similar for the two techniques in the screening environment, but that FFDM was more accurate in women with radiologically dense breasts, including those under 50 and pre- or peri- menopausal women [31].

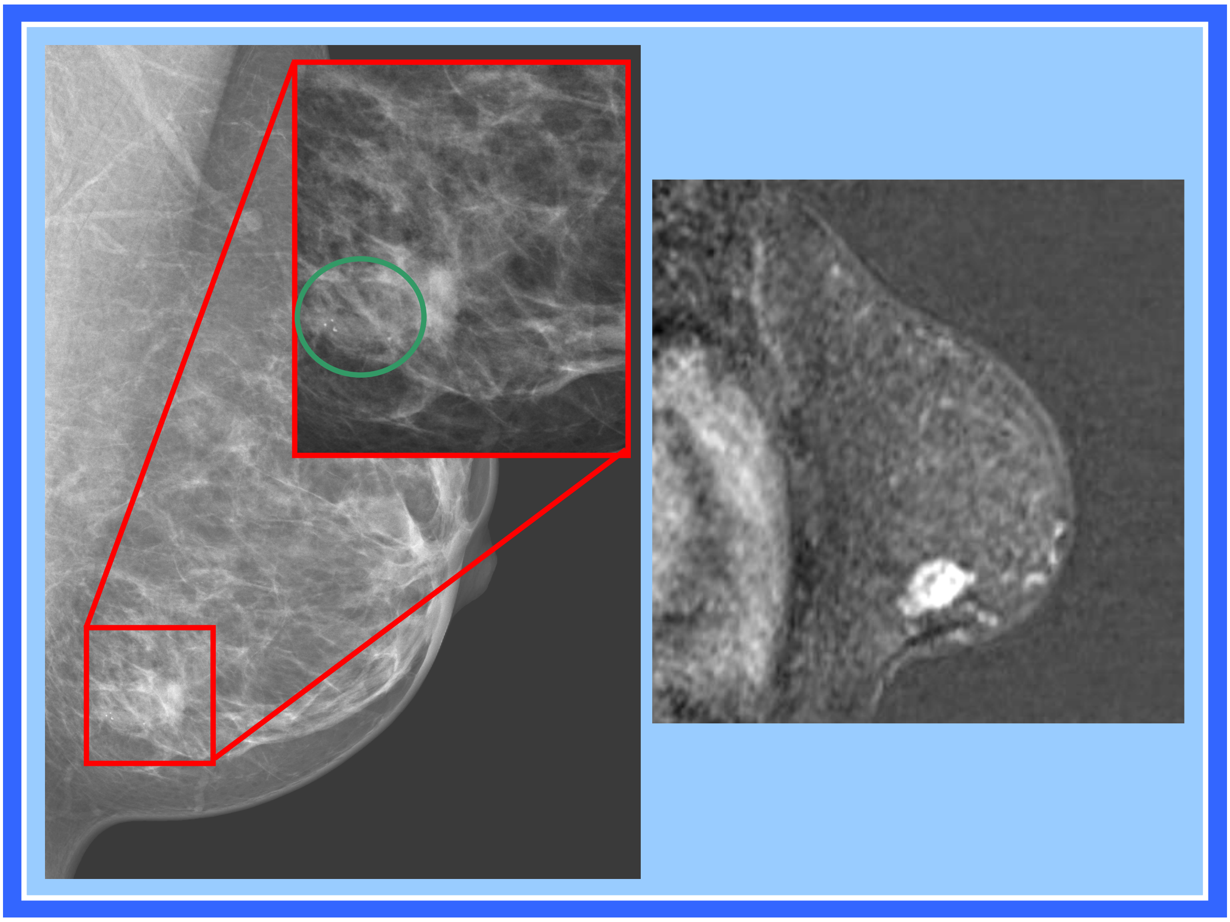


Figure 1.5- Left hand image shows mammography of a grade 3 infiltrating lobular cancer in the left breast. The enlarged insert view on the left hand side shows calcification within the lesion. The right hand image is a contrast enhanced magnetic resonance imaging view demonstrating the same cancerous lesion

As the detection of malignancy using mammography is reliant upon identification of regions with differing x-ray attenuation properties, cancers with similar densities to normal parenchymal tissue, and many of the BRCA gene mutation cancers, will appear mammographically benign [32, 33]. However, the major advantage of mammography is that it is a relatively cheap and quick screening tool and can identify microcalcifications within the breast. Whilst this is not directly indicative of cancer, specific distributions can be suggestive of malignancy [34, 35] as calcium appears to be linked with necrosis associated with breast cancer [36].

One of the main limiting factors associated with mammography is that the sensitivity of the modality is inversely proportional to the density of the parenchymal tissue. Reports suggest that the sensitivity drops from 87% for fatty breast tissue to around 45-63% in dense fibroglandular tissue [26, 27, 37]. As dense breast tissue is known to be a risk factor in breast cancer development, this results in a confounding problem that women with higher risk of developing breast cancer could have the lowest mammographic sensitivity [38].

Mammography uses low doses of ionising radiation, which is a known risk factor in cancer induction [15]. One group have suggested that to prevent inducing more cancers than it leads to treatment for, mammographic screening should only be performed in women over the age of 40 [38].

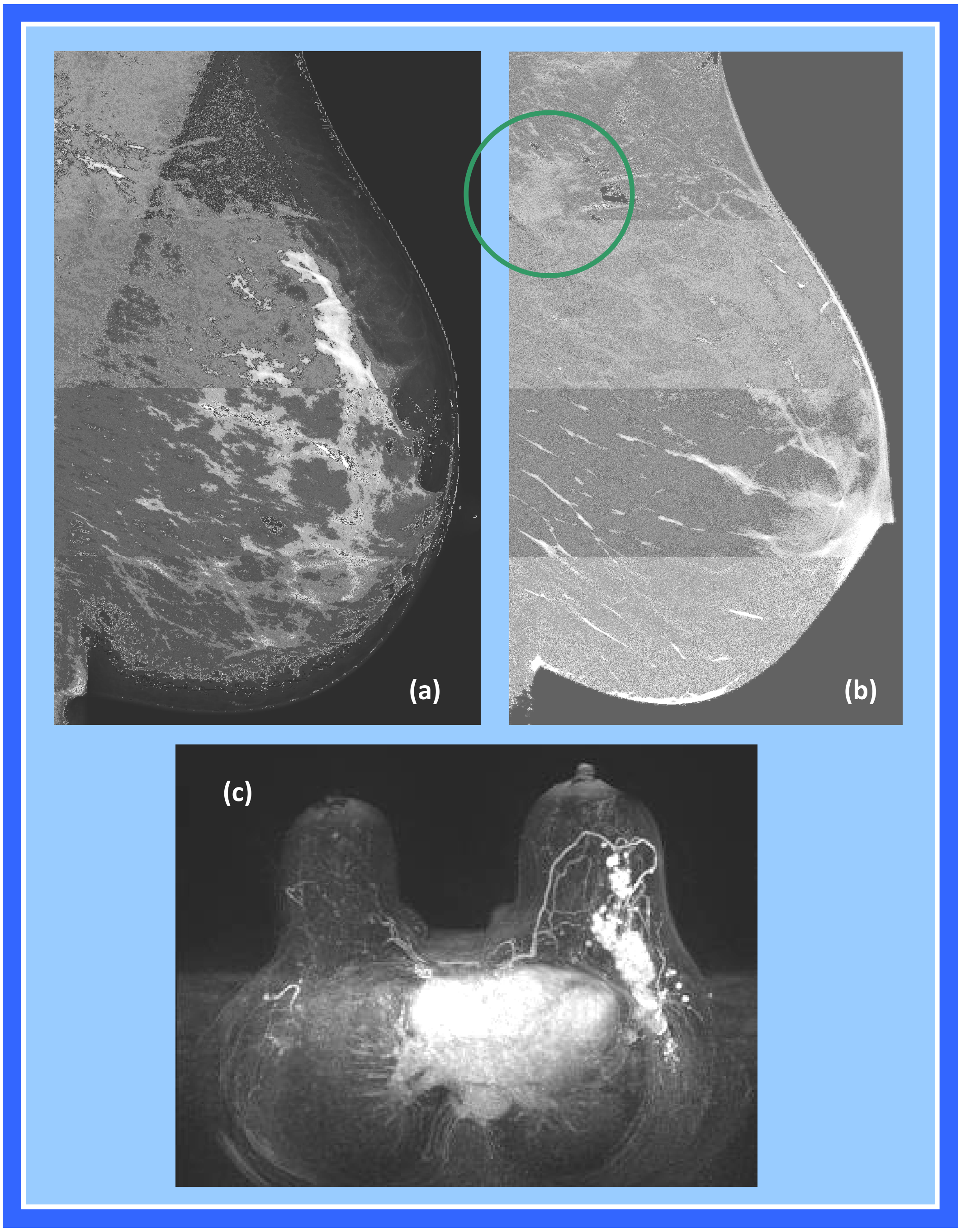


Figure 1.6- (a) conventional digital mammography image from a large ductal grade 3 cancer extending into the left axilla (b) shows DBT image, with cancer clearly demonstrated as indicated (c) corresponding contrast enhanced magnetic resonance imaging showing extent of lesion

Recent advances in x-ray based mammography have been towards breast tomosynthesis, or DBT, which acquires a number of projection images at various angles around the breast. By moving the x-ray source in an arc around the breast, these projection images can be reconstructed in order to produce high resolution slices (Figure 1.6). Doses are reported to be similar, while providing an improvement in contrast detail and a significant improvement in diagnostic accuracy over conventional 2D mammography [39].

1.3.3 Ultrasound

Ultrasonography (US) uses a transducer to transmit high frequency sound waves through tissue and forms the final image using information extracted from the reflected waves from muscle, fluid and solid masses. Lesions are identified by comparing normal breast parenchyma with those of suspicious regions [33] and US is particularly useful in differentiating between solid masses and fluid filled cysts [40]. The blood flow to areas of suspicion can be mapped using Doppler imaging (a technique measuring frequency shifts of sound waves to determine speed and direction of moving structures, such as blood) and this can also provide further information as to whether a lesion is benign or malignant [40].

Ultrasound is relatively independent of breast density and therefore is useful in younger women and those with dense fibroglandular tissue, however it can be very operator dependent [32] and is a time-consuming examination. It can be problematic to identify deep abnormalities and microcalcification cannot always be reliably identified due to the inherent ‘speckle’ of US imaging [40, 41].

1.4 BREAST MAGNETIC RESONANCE IMAGING

Breast Magnetic Resonance Imaging (MRI) is a specific application of MRI utilising a dedicated breast radiofrequency coil (Figure 1.7) that encompasses both breasts to provide high resolution anatomical imaging that is not degraded by dense fibroglandular tissue, scar tissue, radiotherapy changes or breast implants [42]. It has become more routinely used in recent years in the context of problem solving, monitoring treatment response, implant assessment and in the screening of women classified at high risk of developing breast cancer.

In a study comparing the accuracy of the conventional triple assessment to MRI, it was found that the sensitivity of mammography, ultrasound and physical examination combined was still lower than that of a Breast MRI examination (93.2% vs 94.4%) [27]. One of the main advantages of the MRI examination is the high negative predictive value, and therefore a negative study can almost always be considered conclusive proof of no malignancy (Table 1.4).

Figure 1.7- Image of the Siemens dedicated 2-channel breast matrix coil

	Sensitivity	*Specificity*	*PPV*	*NPV*	***Accuracy***
MRI	91.0-94.4 %	26.0-88.0 %	66.0-73.6%	91.7-100.0%	72.9-92.2%

Table 1.4- Reported sensitivity, specificity, positive and negative predictive values and accuracy of breast MRI examinations as reported in the literature [27, 43-46]

The breast MRI examination utilises different techniques in order to provide information on anatomy and physiology of the tumour, which is why one examination provides so much information. The theory of MRI is discussed in detail in Chapter 2, however a brief outline of the key features of the MRI examination is provided below.

1.4.1 Morphological Imaging

MRI images hydrogen nuclei, which are found in the water of which the body is mainly comprised. By imaging this water, excellent soft tissue contrast is realised and the technique is relatively unaffected by fibroglandular density, making it particularly suited for screening and imaging younger women where exposure to ionising radiation is less favourable and breast tissue density is generally higher.

High resolution morphological imaging generally provides an in-plane pixel size around 1×1mm or less and thus tumour morphology can be classified. Sequences that highlight water content (known as 'T_2 sequences') can be useful in identification of intramammary cysts, classification of lymph nodes as benign or malignant by clear visualisation of the intracapsular fat, and finally as a useful adjunct to other sequences in the differentiation between benign and malignant lesions. Kuhl et al. reported that these sequences can be a useful tool in interpretation of whether a mass may be malignant or not as many benign tumours demonstrate a rich extracellular oedematous matrix which appears as increased signal on such T_2 sequences [47]. The majority of malignant lesions behave similar to normal breast parenchyma, with a dense cellularity and therefore will appear as low signal on such imaging sequences [47].

1.4.2 Kinetics

The advantage of Breast MRI is that physiological information regarding tumour vascularity can be obtained. An imaging sequence that is sensitive to the presence of contrast agent is used (known as a 'T_1 sequence', described further in Chapter 2) such that areas that take up contrast are highlighted by bright signal. By imaging both breasts rapidly as contrast is injected, contrast uptake can be visualised dynamically, producing a representative impression of tumour vascularity and potentially angiogenesis [42].

As angiogenic vessels are formed rapidly in order to facilitate tumour growth, they tend to be permeable with large endothelial fenestrations. The contrast is rapidly taken up in areas with an abundance of these vessels due to the increased microvascular density, which leads to an early and strong post contrast enhancement. Subsequently, the contrast leaks into the extra-cellular space, giving rise to a 'Type III' wash-out signal intensity time curve as demonstrated in Figure 1.8 [48, 49]. As normal regions of fibroglandular tissue will display a continuous steady uptake of contrast agent ('Type I'- Figure 1.8), areas of strong early enhancement are readily identifiable. It is this physiological process that makes Breast MRI a particularly sensitive imaging modality, with reported sensitivities of 100% [43, 46, 50].

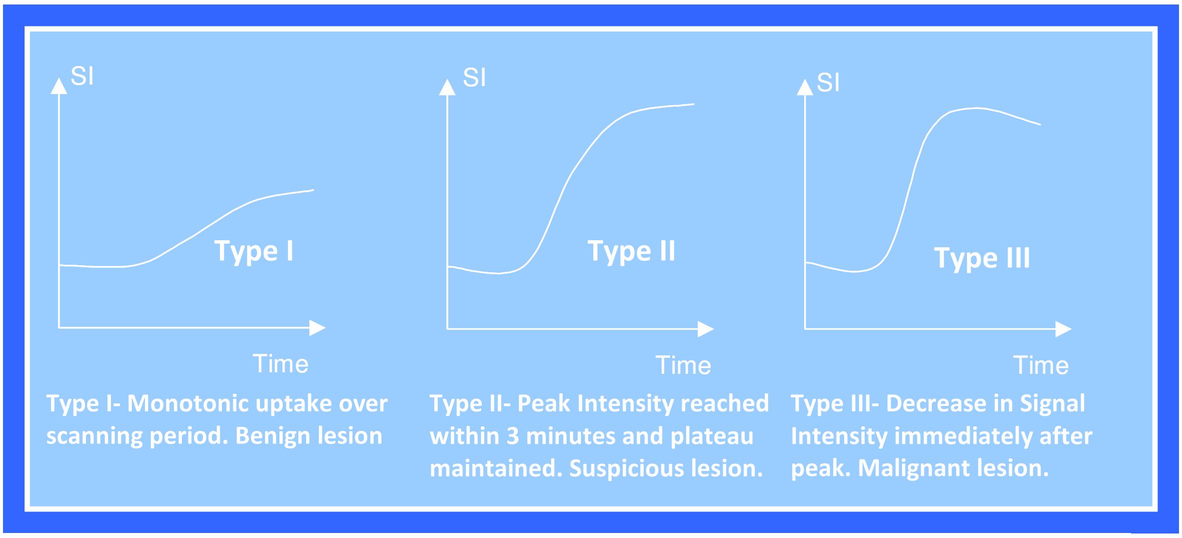


Figure 1.8- Uptake and washout patterns after administration of intravenous contrast

Whilst there is lower neoangiogenesis associated with intraductal tumours, MRI has been able to reliably identify and diagnose DCIS, particularly high grades, due to the greater vessel density associated with such areas [16, 51].

1.4.3 Lesion Diagnosis on MRI

The Breast Imaging Reporting and Data System (BI-RADS) reporting method has been widely used in mammography and breast ultrasound in order to standardise reporting of imaging examinations and provide categories that are widely recognised. In 2003 this BI-RADS lexicon was extended to include standardised reporting of breast MRI examinations, as described by the

American College of Radiology [52]. In order to obtain a BI-RADS category, both morphological and kinetic data is described.

Morphology is described in terms of shape, margin and the internal enhancement characteristics. Spiculated margins and peripheral enhancement of masses is most commonly associated with malignant breast lesions [53]. The distribution and pattern of non-mass like enhancements are also dealt with as well as the assessment of the kinetic curves.

Final assessment categories are defined from BI-RADS:1 (no lesion identified and return to routine follow-up) up to BI-RADS:6 (known malignancy requiring appropriate action) which allow standardised reporting methods across all imaging modalities used in diagnosis and staging of breast cancer.

Breast MRI has demonstrated very high sensitivity in identification of malignancy within the breast (Table 1.4) and while the specificity has been reported as variable in the literature [27, 43-46], it is generally accepted to be around 90% for expert radiologist readers [45]. There are a range of factors that can affect the uptake of contrast within the breast and it has been reported that there is an increase in unnecessary biopsies and other investigations, based on the findings from MRI. One report states MRI overestimates the size of identified malignant lesions [27], while another reports that the use of the examination has led to an increase in avoidable mastectomies, with no proven improvement in the future re-operation rate [45, 54].

Due to the hormonal influences on breast tissue, the stage of the menstrual cycle can influence the contrast uptake within normal, healthy breast parenchyma. During the luteal phase of the cycle (days 20-28), normal healthy fibroglandular tissue undergoes increased proliferation [55, 56], which can in turn lead to foci of enhancement that exhibit suspicious contrast uptake characteristics. Generally, by imaging patients when the hormonal influences are minimal (usually on day 7-13 of the menstrual cycle), this effect can be minimised [57], although this is not always an option in patients with a known cancer.

There are similarly a number of benign changes within the breast that can appear suspicious on MR imaging, particularly in the case of a fibroadenoma, formed from lumps of glandular tissue within the breast. Despite the benign morphological appearance (usually rounded lesions with definite margins), contrast uptake can result in a highly suspicious appearance due to the high vascular density- often similar to that of invasive breast cancer [58].

Other benign changes within the breast that cause neo-vascular changes also require careful classification- for example wound healing or inflammation can result in increased and suspicious uptake patterns within such regions [42].

1.4.5 Advanced MRI Techniques

A number of advanced imaging techniques are becoming more widely utilised within all aspects of MRI, and particularly within cancer imaging. The techniques that are discussed briefly in this section are those that are becoming more clinically applicable.

Assessment of tumour vasculature is assessed by rapidly acquiring the breast volume during contrast acquisition in order to model the washout parameters and use this to infer malignancy as outlined above. However, there have been methods described that model the upslope of the contrast wash-in and use this to identify malignancy, in a technique known as pharmacokinetic modelling [59, 60]. By utilising acquisition times below 20s and fitting contrast parameters to a pre-defined compartmental model, parameters can be fitted to estimate blood and extracellular volumes as well as transfer rates between each compartment [60, 61]. Such techniques, however, often compromise the spatial resolution, which is critical for assessing morphology in order to fully characterise the lesion in terms of the BI-RADS lexicon. Recent drives towards high spatiotemporal sequences have produced promising results [61], however the sophisticated computational analysis techniques have slowed the progression of this technique from the research arena into routine clinical practice.

Diffusion imaging uses a special imaging sequence in order to characterise the mobility of water molecules [62, 63]. As malignant tumours consist of densely packed cells, the diffusion of water molecules is generally restricted within these areas and therefore imaging sequences sensitised to this can provide information relating to the cellularity and therefore the tumour grade [64]. Quantitative values of diffusion coefficients can be calculated and these have been reported to reliably differentiate between benign and malignant lesions [64-68] with malignant lesions generally exhibiting lower diffusion coefficients relative to surrounding tissues, however absolute cut-off values cannot be universally applied due to subtle differences in sequence acquisition parameters [67] and there can be overlap in diffusion characteristics between invasive and non-invasive cancers. Despite this, however, diffusion has been reported to increase the specificity of breast MRI to around 81-88% [66, 68]. Benign lesions exhibiting high cellularity such as papillomas can appear malignant on diffusion imaging and therefore diffusion imaging cannot be used alone for the diagnosis of breast malignancy.

Other, more specialist, techniques such as spectroscopy [69, 70] and magnetisation transfer imaging [71] have also been demonstrated to potentially increase the specificity of the examination and shown initial promise in identification and classification of malignancy. However, these techniques are still very much research based and are not widely available and therefore have yet to demonstrate their full potential.

1.5 IMPROVING LESION DETECTION AND CLASSIFICATION

1.5.1 Importance of Morphological Classification

The majority of the BI-RADS reporting criteria focuses on morphology as this is recognised to be an important factor in identification of malignancy. Assessment of this technique, however, can be subjective and dependent on the reader experience.

It has been reported that radiologists differ substantially in the interpretation of images using the BI-RADS classification. One group report that when 14 observers reported ultrasound images of known breast lesions using the BI-RADS system, the agreement using the kappa statistic (κ) was 0.39 (kappa is a measure of agreement with categories less than 0.21, 0.21-0.4, 0.41-0.6, 0.61-0.8 and 0.81-1.0 considered poor, fair, moderate, good and very good agreement respectively) [72]. Another group condensed the BI-RADS categories down into 3 categories rather than 5, and with this they estimated $\kappa=0.48$ for ultrasound and $\kappa=0.58$ for mammography [73]. This is in agreement with Redondo et al who reported $\kappa=0.53$ when the BI-RADS categories were categorised into further investigation required (BI-RADS 3-5) or not (BI-RADS 1 and 2), however when all five categories were included, the reported agreement was significantly lower at $\kappa=0.37$. This group also reported intra-observer agreement and found moderate agreement when the full BI-RADS scoring system was used ($\kappa=0.53$) [74]

There are fewer reports on the agreement between readers for breast MRI, although one study found that between expert breast radiologists using a scoring system identifying whether a mass was deemed to be spiculated or not resulted in κ=0.38-0.56 on axial images, and the agreement was significantly lower on sagittal imaging (κ=0.27-0.51) [75].

Due to the reported variability in morphological assessment of lesions on different imaging modalities, a standardised more objective measure is required in order to improve diagnostic accuracy and certainty. There has been a drive over the past decade or so to develop Computer Aided Detection (CAD) systems that will have the capability of identification and classification of lesions on various imaging modalities.

With the advent of digital technology such as digital mammography, development of CAD systems to analyse digital images has become simpler. Although the technique utilised for each imaging modality is slightly different, generally a thresholding step is used to remove false signals and then segmentation performed prior to classification. These are then identified on the image to prompt an experienced user to the region. Whilst some groups report an increase in the sensitivity [76] and specificity [77], one group report that the inclusion of CAD in a large scale study considering over 115,500 mammograms had no statistically significant impact on breast cancer detection or recall rates [78]. Therefore it is unclear whether the inclusion of CAD in a screening environment would have a significant impact on overall breast cancer detection rates, although it is possible it may be a useful tool for junior radiologists [79].

CAD systems for ultrasound have been less well described and are more complex, requiring descriptors of shape, margin and mass characteristics to be quantified. However, Shen et al report a sensitivity and specificity of 90.6% and 92.2% respectively for the CAD system they designed, which computerised the BI-RADS mass features [80]. They describe that the most useful features were the 'angular characteristic' feature which describes the margin and the abruptness of the lesion boundary, both of which provide a measure of how spiculate a lesion is on the image.

CAD systems for MRI are mainly based around analysis of kinetic contrast uptake data [81-83] due to the volume of images generated, however there have been reports on morphological assessment of lesions using in-house and commercial software [84, 85]. These software packages utilised descriptors of shape in order to describe the shape and margins of the lesion, as well as 'texture' features which were used to quantify the internal characteristics of the lesion on the image. This resulted in reported diagnostic accuracies of 93.5% [84], with area under operator receiver characteristic curves (a method of describing accuracy of diagnostic tests, see Chapter 2.3.2) of 0.82 [85].

1.5.2 Texture Analysis

Texture is a difficult concept to define, however is generally described using words such as rough, smooth, dense, random etc. [86]. When applying this to images, the texture is characterised by the spatial distribution of the grey level pixel intensities and by quantifying these, patterns can be identified and classified.

In medical diagnostics, it has been reported that radiologists visually assess texture in clinical images in order to diagnose pathology [87] and therefore if computer software could be used to mimic the expert eye, it could prove useful in identification of abnormality within tissue. This technique is known as 'texture analysis' (TA).

Texture analysis is computationally demanding as the assessment of the distribution of pixel intensity values requires calculation of higher-order statistics [88] as well as fitting to models [89] and rescaling the data in order to obtain maximum information from the images. This technique allows differentiation of regions by considering the image on a pixel-by-pixel basis and therefore can detect changes that may not be visually apparent [90].

Texture analysis has been widely used in medical diagnostics. It has been demonstrated to reliably characterise brain tissue and differentiate between tumour and oedema [87, 91] and has also been used to identify abnormalities that have not been visually identified in epilepsy sufferers [92, 93]. Currently

Alzheimer’s disease is only accurately diagnosed by a histological sample [94], however Nedelec et al. reported that TA may prove useful in the diagnosis and monitoring of therapies for this debilitating disease [92]. Brain gliomas are difficult to differentiate from metastases in the brain; however Zacharaki et al. have demonstrated that using TA, the two can be differentiated with concomitant sensitivity and specificity using MRI images [95].

In other areas of the body, TA has been used for identifying cirrhotic lesions in the liver, and differentiating these from normal healthy liver [96]. It has also proved to be a useful tool in the classification of osteoporotic bone [97, 98].

More recently, the technique has become more widely used in cancer imaging, not only in the identification and classification of disease [99-105], but also as a potential predictor of prognosis for the disease and treatment stratification [106-108].

1.5.3 Texture Analysis in Breast Imaging

Texture analysis has been used extensively in breast imaging. In mammographic images, texture analysis has been used to differentiate between malignant and normal breast masses [99, 101] as well as between cluster patterns of calcification in order to discriminate between invasive cancer and DCIS [109]. Similarly in ultrasound imaging of the breast [100, 102], Garra et al demonstrated that texture analysis could be used to reduce the number of biopsies carried out on benign lesions [102]. In both modalities, texture analysis has been demonstrated to increase the specificity [99, 100, 102] and therefore it is likely it could potentially be used for the same purpose in Breast MRI.

There are an increasing number of journal articles published investigating the use of texture analysis in breast MRI [103-105, 110]. The technique has been demonstrated to distinguish between benign and malignant lesions [103, 104] and a small, preliminary report suggests also between different histological breast cancer types [110].

The main limitation of TA is that reports in the literature suggest that the technique is not transferable between sites or scanners [111]. It has been suggested that data acquired in one centre will not necessarily correlate with that acquired elsewhere. This could potentially be due to differences in sequence acquisitions, RF coil architecture and image processing, which may result in dissimilar levels and structures of noise and received signal across the final image.

1.6 AIMS AND SCOPE OF THESIS

A comprehensive literature review suggests that texture analysis in breast magnetic resonance imaging shows initial promise in being an exciting area for further development. The aim of this thesis is to more thoroughly investigate the role of the technique in a clinical department utilising images from routinely referred patients, recently diagnosed with primary breast cancer.

In order to utilise the technique, an assessment of the robustness of the technique and influences of image acquisition conditions will have to be considered prior to application on routinely acquired images.

The role of texture analysis in the identification, classification and staging of breast cancer in patients referred for breast MR will then be considered, and the research seeks to extend the preliminary work that has already been published by using the technique to classify breast cancers into their respective histological and molecular subtypes. To fully assess the usefulness of the technique in diagnosis, a classification model will be built and tested and an assessment made of the classification accuracy and therefore clinical applicability.

A preliminary investigation into TA as a tool in the neoadjuvant chemotherapy setting will be considered as this is a new, developing, area of MR in the monitoring of such treatment.

This work will hopefully lead to a consideration as to whether a currently research-only tool could potentially be applied in a clinical setting and the role it may play in the management of patients referred to the MR unit through the breast clinic.

CHAPTER TWO: THEORY

2.1 MAGNETIC RESONANCE IMAGING

2.1.1 Hardware- Magnets, Gradients and Radiofrequency System

In order to perform magnetic resonance imaging (MRI), a large magnetic field is required. This magnetic field is typically on the order of 1.5 Tesla (T) for clinical applications to 3.0 T and higher for research applications. In order to generate such large, stable magnetic fields, superconducting magnets are generally used in modern MRI scanner design (see Figure 2.9 [112]). These magnets are manufactured from coils of superconducting wires (typically niobium-titanium) which are held in a copper matrix for mechanical stability. Once cooled below a critical temperature (T_c=7.7K, -265.3°C) the wires exhibit no resistivity and therefore after sufficient current is in the windings to generate the required field, the power supply can be removed and persistent currents will flow with no degradation due to losses from electrical resistance. Modern superconducting

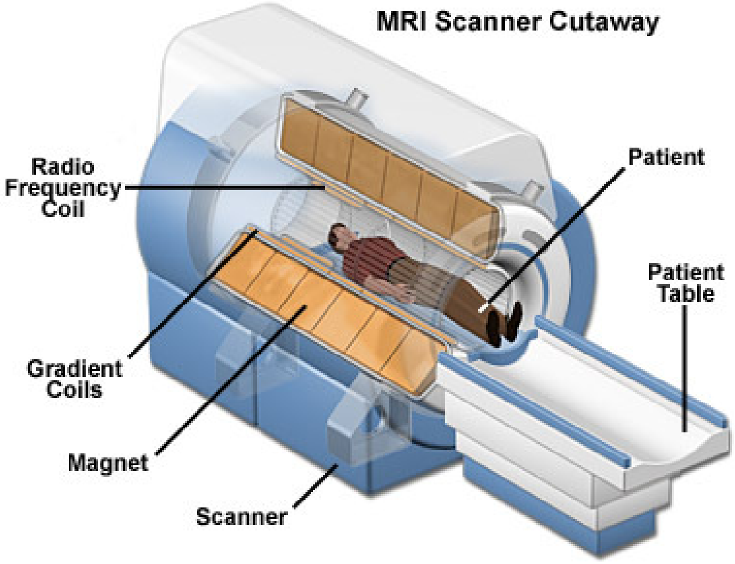


Figure 2.9- Schematic diagram of the inside of MRI scanner main bore housing [111].

magnets have coils of superconducting wire cooled in a bath of liquid helium (boiling point 4.2K, -268°C). A refrigerated cold head is used to ensure minimum helium boil-off.

Gradient coils are used for signal localisation in order to create the final image and are mounted just inside the bore of the magnet and are held at room temperature, with water cooling to remove excess heat. These coils are built into the main scanner and produce smaller magnetic fields that either add or subtract from the main static field, resulting in a linearly varying magnetic fields. Electromagnets are used with amplifiers that control how rapidly gradients can be switched (have the polarity of the current reversed) and the maximum field strength they can produce. Faster and stronger gradients allow for higher image resolution and faster image acquisition times.

The final essential component for creation of MR images is the radiofrequency (RF) system. This comprises of two parts- an RF transmitter and RF receiver. The RF transmit system creates an external oscillating magnetic field at a resonant frequency which is absorbed by the nuclei and disturbs them from their thermal equilibrium position. Once the external RF source is switched off, the nuclei will then return to their

Figure 2.10- InVivo 7-channel breast coil

equilibrium position, and in the process they create an external oscillating magnetic field which can be detected by a receiver coil, due to the induction of a small current. As these currents are very small, receiver coils have to be placed as close to the region being imaged as possible. In MRI, specific coils are manufactured for each imaging application- e.g. head coils, spine coils, breast coils etc. (see Figure 2.10) which are designed to fit closely around the imaged area to maximise the signal received.

Modern MRI systems are controlled by highly sophisticated computer control systems which manage not only timing of gradients, RF transmit/ receive pulses and acquisition processes, but also control safety systems and ensure safe modes of operation.

2.1.2 Basics of Nuclear Magnetic Resonance (NMR) Theory

Nuclei of some atoms possess small magnetic fields which arise due to the nucleus possessing a charge and spinning. The resulting moving net charge results in a magnetic moment. For conventional, routine MRI, hydrogen is the most commonly imaged nucleus due to the plentiful abundance in the human body (more than 70% of the atoms in the body). Due to the large number of randomly distributed individual moments, the net magnetisation is zero, however when placed in a strong external magnetic field, the individual moments can either align parallel to the field (low energy state) or anti-parallel (high energy state). In these energy states, the moments precess around the external field axis at a given frequency- the Larmor frequency, ω_L, which is governed by the gyromagnetic ratio, γ, and the strength of the external field, B_0 (Equation 2.1).

$$\omega_L = \gamma B_0$$

Equation 2.1

The gyromagnetic ratio is a constant for any given nucleus and is the ratio of the magnetic moment and the angular momentum. For hydrogen, γ=42.58 MHz/T. At thermal equilibrium, there will be a small excess of spins in the low energy state relative to the high energy state- around 4 spins per million at 1.5T. We can define a bulk magnetisation vector, **M**, that describes the combined effect of the sample rather than considering the motion of individual spins, which in equilibrium conditions is aligned along B_0 (M_0). By viewing this bulk magnetisation vector, **M**, from a rotating frame of reference also rotating at ω_L, and using classical mechanics, the description of complex excitation and relaxation processes is simplified.

In order to infer information regarding the surroundings of the nuclei of interest, the system has to be perturbed from its equilibrium state. Energy is delivered into the system in the form of an

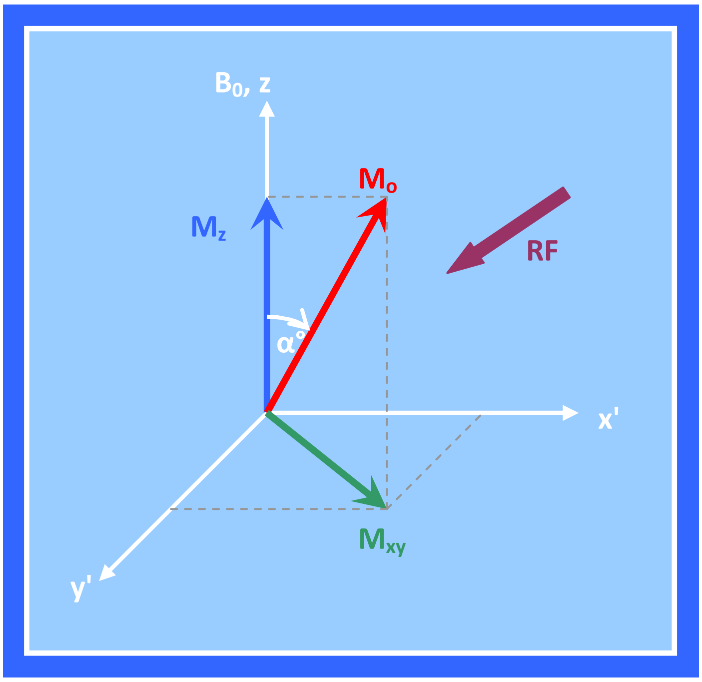


Figure 2.11- Rotating frame depiction of the effect on bulk magnetisation vector, **M**, of an applied external RF pulse

external radiofrequency (RF) field applied perpendicularly to B_0, with a frequency equal to ω_L. This external field is conventionally denoted as B_1. This RF field has the effect of 'tipping' **M** away from B_0 in a spiral trajectory (a simple arc in the rotating frame), and the angle which it makes to the M_0 position is called the flip angle, α, which is dependent on the strength and time that B_1 is applied for (see Figure 2.11).

The component of **M** perpendicular to the main static field (M_{xy}, see Figure 2.11) generates a current that can be measured, and it has a maximum value when there is no longitudinal (M_z) component- i.e. in the transverse plane of the rotating frame of reference.

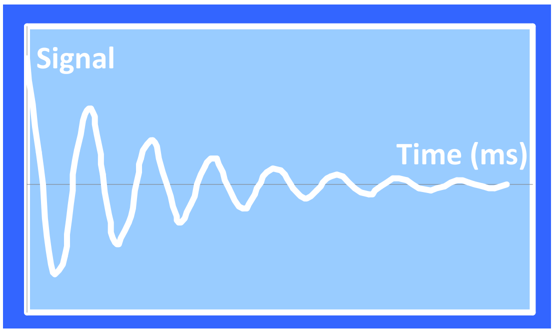


Figure 2.12- Free induction decay signal

Once the B_1 field is removed, the signal rapidly decays to zero as the protons undergo relaxation processes in order to return to thermal equilibrium. The resulting Free Induction Decay (FID) signal, Figure 2.12, contains information relating to the environment from which the signal was received and this is the NMR signal that is used for producing images in MRI.

2.1.3 Relaxation

Relaxation is the process by which **M** returns to the equilibrium condition, M_0. There are two processes of relaxation in NMR: longitudinal and transverse relaxation.

Longitudinal relaxation is also known as spin-lattice, or T_1, relaxation. It is progressive loss of thermal energy to surroundings to realign magnetic moments with B_0. The time for this process to occur is governed by the microenvironment and occurs with a time constant T_1, which is the time for 63% of **M** to return to M_0 (Equation 2.2).

$$M_z = M_0\left[1 - \exp\left(-\frac{t}{T1}\right)\right]$$

Equation 2.2

Transverse relaxation, also known as spin-spin, or T_2, relaxation describes the loss of phase coherence between the nuclei. As each individual precessing moment produces its own local magnetic field, neighbouring spins experience a slight perturbation of the main field and precess at slightly different frequencies around $\omega_{L,}$ therefore there is a loss of phase coherence and the transverse component of the magnetisation vector, M_{xy}. The time constant associated with this, T_2, is the time for M_{xy} to be reduced by 63% (Equation 2.3).

$$M_{xy} = M_0\left[\exp\left(-\frac{t}{T2}\right)\right]$$

Equation 2.3

Inhomogeneities in B_0 also result in small changes in precessional frequencies that cause an additional dephasing component that result in an effective time constant T_2^*, which is shorter than T_2 (Equation 2.4).

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \gamma \Delta B_0$$

Equation 2.4

These relaxation properties are directly related to the environment and associated movement of the molecule containing the excited spins. The movement of the molecules causes slight fluctuations in the Larmor frequency due to changes in magnetic field, and the characteristic time of these interactions between molecules is described by the correlation time, τ_c. Free water has short correlation times, due to rapid, free movement of the water molecules, while bound water (for example water molecules bound to proteins) exhibits long correlation times.

Due to the short correlation times of free water, the rapidly oscillating magnetic field tends to result in a resonance slightly higher than the Larmor frequency, while solids exhibit a slightly lower precessional frequency.

In order to have efficient T_1 relaxation, the resonance condition of $\omega=\omega_L$ is required, and where this is not met, T_1 times are longer due to inefficient transfer of energy to surrounding molecules- see Figure 2.13.

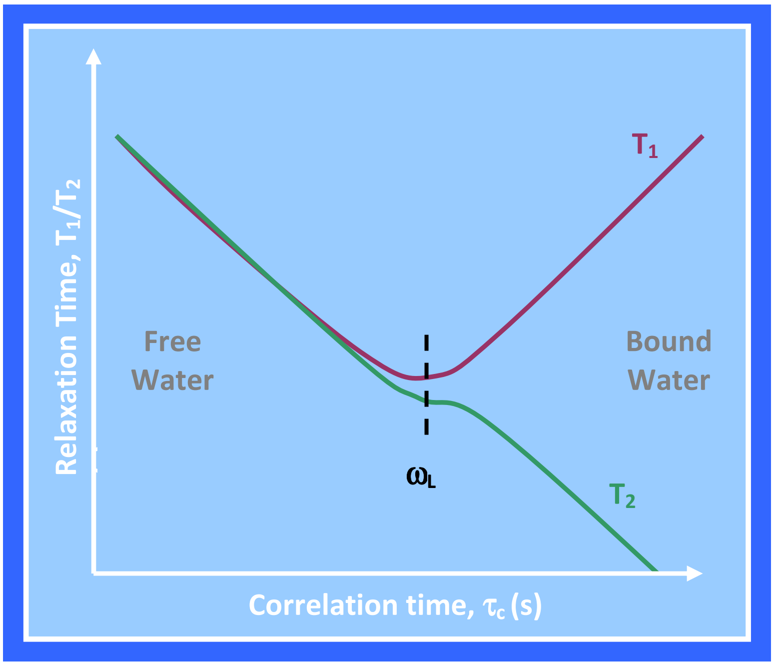


Figure 2.13- Relationship between correlation times, τ_c and T_1 and T_2 relaxation times

T_2 relaxation times are also affected by τ_c. Where τ_c values are short, such as in free water, interaction times are short and therefore the rate of change of phase is lower than compared with bound water, resulting in long T_2 relaxation times (Figure 2.13). In the case of bound water, τ_c times are longer resulting in a more rapid loss of phase coherence between the molecules and shorter T_2 relaxation times as protons experience both local fields from bound water and applied external fields. As the T_1 relaxation is relatively unaffected by low frequency fluctuations, longitudinal relaxation times are always longer than transverse relaxation times.

2.1.4 NMR Signal Localisation

In order to localise where the NMR signal has been received from, it is necessary to introduce a known relationship between precessional frequency, phase and spatial position. This is achieved by using magnetic field gradients, which are additional, smaller, magnetic fields applied either during RF excitation, between excitation and signal detection, or during signal detection.

The slice selection gradient is applied during excitation to ensure that only a thin section of the patient experiences the excitation RF pulse, thus ensuring that signal from only one slice of the patient is received. The gradient is applied perpendicular to the required slice orientation, and results in the frequency of precession varying linearly along this axis. Due to the resonant nature of NMR, energy will only be transferred to spins precessing at ω_L and only these nuclei will produce an FID. The slice width is determined by the RF bandwidth and the gradient strength.

During signal detection, a gradient is applied in one direction across the slice to encode the received signal, such that the received frequency will be directly related to the position within the slice. This is known as the frequency encode gradient and results in a complex signal relating to the position and strength of the signal received across the slice.

Phase encoding is used for spatial encoding in the final orientation of the slice. A gradient is applied for a short time between RF excitation and detection. The gradient causes relative phase shifts across the plane of the slice due to small changes in precessional frequencies caused by the gradient. These phase shifts are retained after the gradient is switched off, when spins return to precessing at the Larmor frequency. The detected phase change can be directly linked to the spatial position via the gradient. To fully encode the imaging slice, the phase encode step is repeated a number of times with different gradient strengths, requiring a number of excitations. The number of phase encode steps required is defined by the imaging matrix (i.e. for a 256×256 matrix, there are 256 phase encoding steps and therefore 256 excitations).

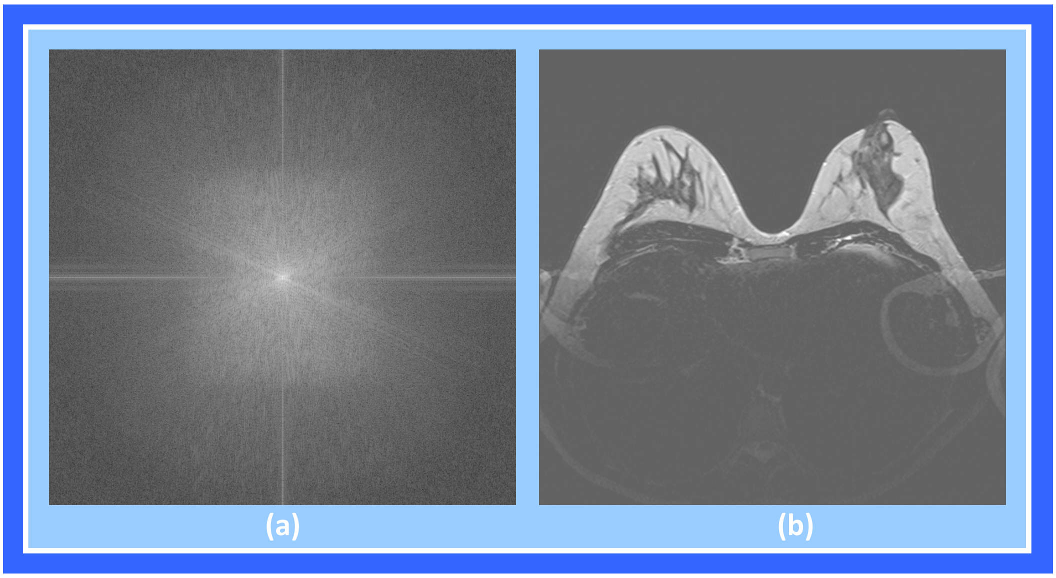


Figure 2.14- (a) k-space raw data from a T2 Spin Echo sequence of the breast
(b) Final image formed by fast Fourier transform of raw data in (a)

The NMR signal is composed of combinations of sine and cosine waves with different frequencies and phases. These complex waveforms contain the information about the area being imaged and are digitised and stored in raw data space. The digitisation of these received signals results in raw data space, which is conventionally known as ‘k-space’ as shown in Figure 2.14. Each data point represents a wave in the final image with the k-space numerical value encoding the final amplitude (brightness), and the position within k-space determining the wave frequency and angle. Therefore every point in k-space is represented in every voxel in the final image. As the values close to the centre of k-space represent waves with longer wavelengths, these primarily represent contrast detail in the final image. Those points on the periphery of k-space have high frequencies and therefore contain the detail of the image.

In order to form the final image from k-space, the constituent frequencies and amplitudes are calculated using the Fast Fourier Transform (FFT). This extracts all frequency information from the encoded signal, and as the gradients imparted a known positional relationship between frequency, phase and spatial position, the NMR signal can be translated to a final image.

2.1.5 Magnetic Resonance Imaging (MRI) - Pulse Sequences

The nucleus of hydrogen contains only one proton and as they are naturally abundant atoms in fat and water, they are used in most MR imaging, although other atoms such as phosphorus (^{31}P), carbon (^{13}C) and sodium (^{23}Na) can be used.

By manipulating the timing, strength and duration of the B_1 field, as well as timing of the signal detection, tissues can be differentiated due to different micro-environments of the hydrogen nuclei. The timings can be described by pulse sequence diagrams.

There are three common sequence types, upon which all others are based. These are Inversion Recovery (IR) sequences, Spin Echo sequences (SE) and Gradient Echo sequences (GE). This thesis will deal with only GE sequences.

2.1.6 Gradient Echo Sequences

Rather than sampling the FID produced immediately after the application of a B_1 excitation, it is more common in MRI to form echoes due to limits on the time to ramp-up gradients for signal readout. In GE sequences, the echo is formed by using the frequency encoding gradient to dephase and rephase the protons to form an echo that can be fully sampled. The rephasing is achieved by reversing the polarity of the dephasing gradient and at the point when the rephasing lobe equals the area of the dephasing lobe, the maximum echo signal is received. The time between the excitation pulse and the maximum echo signal is known as the echo time, TE.

As described earlier, the phase encode gradient has to be applied a number of times to fully spatially encode the imaging slice, and therefore a number of RF excitations are required. The time between each excitation is called the repetition, TR. To prevent saturation of protons, a TR greater than five times the maximum T_1 of the tissue being imaged is required, to allow full longitudinal relaxation prior to the next excitation. In gradient echo imaging, however, TR values are usually significantly shorter than this to speed up the acquisition process. To reduce the effect of proton saturation, flip angles of less than 90° are used for excitation. The contrast in the images is determined by the T_1, T_2,

T_2^* and proton density (PD) of the imaged tissues, and is highlighted by manipulating the parameters associated with the imaging sequence- the flip angle (α), TR and TE.

2.1.7 Fast Gradient Echo Sequences

In order to acquire images faster, it is possible to shorten the TR to reduce the overall scan time. When full T_1 recovery is not possible within each TR, saturation can occur resulting in little or no signal from tissues with long T_1 relaxation times and therefore to reduce this effect, small flip angles are used. Whilst this results in an overall lower signal by not allowing full T_1 recovery to the equilibrium position, a component of magnetisation is retained along the z-axis for further excitations which prevents saturation. For a tissue with a given T_1 value, the maximum signal is obtained by using a flip angle known as 'the Ernst Angle', α_E, as given in Equation 5.5.

$$\alpha_E = \cos^{-1}\left(e^{-TR/T_1}\right)$$

Equation 5.5

Image contrast in fast gradient echo sequences is mainly controlled by the flip angle and TE. To produce T_1 weighted images, it is necessary to ensure that full T_1 recovery has not taken place when the FID is sampled and therefore large flip angles (typically more than 50°[113]) and short TE values (less than 15ms) are utilised. In T_1 weighted images, fluid is dark and fat very bright and such images are most commonly used for looking at anatomy.

To produce T_2^* weighted images, full T_1 recovery is required to remove T_1 weighting and low flip angles (less than 40°) and long TE (>30ms) values are used. T_2^* images show bright fluid, while fat is darker than in a T_1 image and are most commonly used for highlighting pathology.

Proton Density (PD) weighted images provide a representation of the density of hydrogen nuclei in the tissues. This is achieved by removing all T_1 and T_2 weighting in the image (i.e. small flip angle and short TE). These images have less contrast than either T_1 or T_2 as all tissues have similar water content.

Fast GE sequences allow short TR values to be used, thus reducing the overall acquisition times. When TR values are reduced sufficiently that full transverse relaxation cannot occur, remnant transverse magnetisation from previous excitations can cause artefacts on the final image due to the formation of stimulated echoes. There are various methods of dealing with this remnant transverse magnetisation to prevent such imaging artefacts- by either using the echoes to form the final image, combining the echo with the FID signal or by using only the FID signal. This thesis deals only with the final method, in which the remnant transverse magnetisation is destroyed to prevent interactions between successive repetitions. Such sequences are known as spoiled gradient echo sequences, and the Siemens nomenclature for this is 'Fast Low Angle SHot', or FLASH, sequence. FLASH sequences utilise strong 'spoiler gradients' which destroy phase coherence between successive TR intervals. These sequences can provide T_1 and T_2^* weighting, although in the context of this thesis, and Breast MRI, they are used for rapid three dimensional T_1-weighted scanning.

2.1.8 Additional Contrast

As well as traditional T_1, T_2, T_2^* and PD weighted images, further contrast can be introduced into images either by the use of saturation techniques or by the artificial addition of a contrast agent.

Due to the different micro-environments of the hydrogen nuclei in fat and water, the Larmor frequency of each is slightly different- around 220Hz difference in precessional frequencies at 1.5T. In a frequency spectrum there is therefore a slight difference in the fat and water peaks. By applying a pre-excitation saturation RF pulse at the frequency of the fat protons, then immediately applying the excitation B_1 at the ω_L of the water protons, there is no signal received from fat and therefore it appears dark in the final images. Such a technique is known as fat suppression and is commonly used in fatty tissues (such as the breast), as the high signal intensity can mask underlying pathologies.

Contrast agents can also be used in MRI. The most commonly used agents are based on chelated gadolinium-based compounds. As gadolinium is highly paramagnetic, it disturbs the local magnetic field, thus reducing the T_1 in the immediate vicinity. Therefore on a T_1-weighted image, areas of contrast uptake appear bright.

2.2 TEXTURE ANALYSIS

2.2.1 What is Texture?

Texture is a particularly difficult physical property to define, despite the fact it is a concept that is widely identified with. The Oxford English Dictionary defines texture as *''the constitution, structure or substance of anything with regard to its constituent or formative elements'* [114].

The eye processes and identifies the texture of an object by the variance in the light reflected from a surface [115] and therefore it is a physical property relating to the composition of the object in question. The texture is composed from the randomness, periodicity, directionality and orientation of the composite elements making up the structure of the object [115]. The eye can extract complex information from images in order to infer texture properties of the objects in question and therefore much work has been carried out to understand the human perception of texture using computer analysis techniques [88, 116].

The concept of texture becomes more difficult when trying to describe images, but the generally accepted description of texture in an image is based on that which Haralick described in the 1970's. He described texture as comprising spatial distributions of tonal compositional elements [86, 88, 115, 117]- i.e. the spatial distribution of grey-levels within an image.

There are two main areas of texture analysis in imaging- in the classification of different regions within an image and in segmentation of textures in order to establish boundaries between regions [118]. This thesis deals only with the former application.

The visual system can distinguish between regions in an image by using cues such as changes in texture, colours, luminance etc, and in general classification accuracy is increased for more than one visual cue, rather than relying on one stimulus alone [119]. However, as more textures are introduced into an image, it becomes more difficult for the brain to readily classify the data quickly with minimal effort [116]. In such cases, computer methods for describing texture can prove useful in complex discriminatory tasks. There are many ways in which to perform texture analysis. The three main methods for computationally describing texture features as used in this thesis- namely statistical methods, model-based methods and transform methods- are described in the following sections.

Mathematical derivations of features associated with each model can be found in Appendix A.

2.2.2 Statistical Methods for Texture Analysis

Texture can be described using words such as uniform, rough, smooth, directional, random etc. which relate to the pixel values and distributions that comprise the image [86, 120]. This leads us to a statistical description of texture. Statistical methods are classified as first, second or higher-order methods. Statistical methods use the grey level distribution of pixels in the image for texture analysis. Generally statistical based methods will provide better discrimination between classes than structural or transform methods [121] and are the most widely used in medical applications.

Histogram Methods

First order techniques utilise the frequency of grey-levels contained in the image histogram to infer texture properties such as image mean, variance, skewness and kurtosis. Such descriptors are simple, however they do not provide a high discriminative power as no consideration is made of correlation or co-occurrences of more than one pixel [118].

Max-Min Method

The max-min method of texture analysis utilises the visual perception of texture being dependent on the frequency of extremes in grey level intensity. By smoothing the data and determining the number of grey-level minimum and maximum values in a given direction, a texture feature can be defined [122].

Smoothing is carried out on the original data-points (x_k) with reference to a threshold value (T), such that a new smoothed curve is defined (y_k) (Equation 2.6, Figure 2.15).

$$\text{If} \quad y_k < x_{k+1} - T/2 \qquad \text{then} \quad y_{k+1} = x_{k+1} - T/2$$
$$\text{If} \quad x_{k+1} - T/2 < y_k < x_{k+1} + T/2 \qquad \text{then} \quad y_{k+1} = y_k$$
$$\text{If} \quad y_k > x_{k+1} + T/2 \qquad \text{then} \quad y_{k+1} = x_{k+1} + T/2$$

Equation 2.6

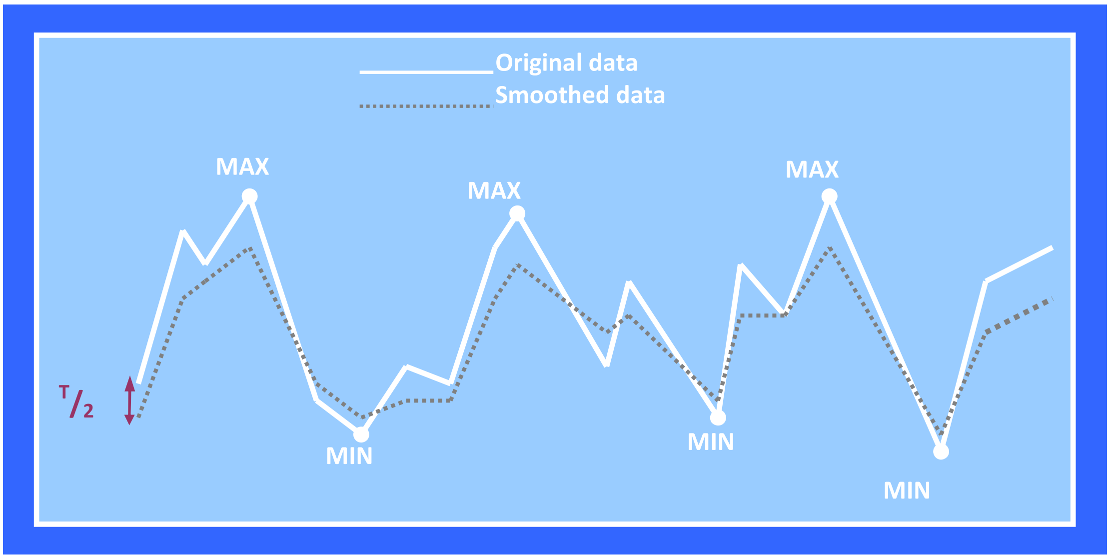


Figure 2.15- Schematic diagram of max-min smoothing and identifications of extreme values

The number of maximum and minimum intensity extremes detected will therefore depend on the threshold value chosen- large T values result in a reduction in max-min values, small T values result in less of a reduction. To fully characterise an image, therefore, a number of T values are used and each time the number of extreme values is calculated. The ratio of extremes at one threshold value to another are then calculated to reduce the dependency on the absolute number of extreme values. These ratio values form the texture description.

Co-Occurrence Matrix

The co-occurrence matrix (COM) is commonly referred to as a 'second-order histogram' as it is concerned with pairs of pixels rather than individual pixels [86]. It represents the count of pixels in a given direction and at a specified distance with co-occurring pixel values of *i* and *j* [88, 121]. A matrix is produced in each direction, θ, for each interpixel distance, *d*, with the matrix dimensions being equal to the number of intensity levels. This can, therefore, become computationally intense and the number of grey levels in an image would normally undergo a rescaling and re-binning procedure to reduce the range of pixel values contained within an image. While this process increases the counting statistics of the COM, it leads to a potential reduction in the discriminatory power of the model.

There are eleven texture features derived from the co-occurrence matrix that will be used in this work, calculated in four directions and for distances up to five pixels. The full list of statistical texture features is listed in Table 2.5.

Images with identical second order statistics are visually indistinguishable [123] and therefore computer-aided texture analysis is essential for assessing texture on a pixel-by-pixel basis, and for allowing potential discrimination.

Gradient Features

The gradient model considers the relationship of variations in grey level intensities across neighbouring pixels. High gradients are produced when pixels change intensity rapidly, while low gradients are produced for gradually varying pixel intensities [121]- see Figure 2.16 below.

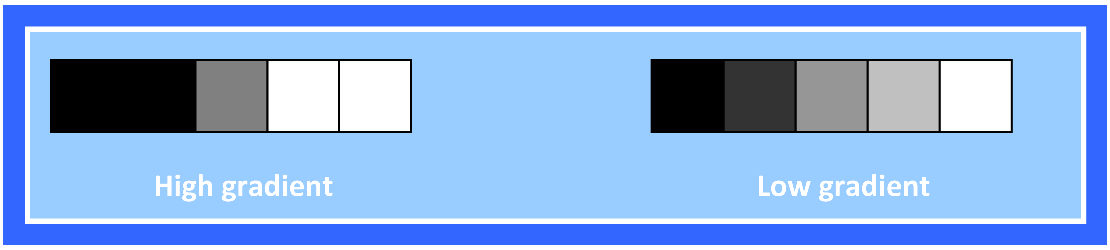


Figure 2.16- Pictorial depiction of high and low gradient in an image when considering the three central pixel grey levels

Texture features such as the mean gradient, kurtosis and variance can be calculated as shown in Table 2.5.

			Number of Features calculated
Co-occurrence Matrix	COM	angular second moment, contrast, correlation, difference entropy, difference variance, entropy, inverse difference moment, sum average, sum entropy, sum of squares, sum variance *(in four directions and up to interpixel distances of 5)*	220
Absolute Gradient	GRA	gradient mean, kurtosis, nonzero, skewness, variance	5
Run Length Matrix	RLM	fraction of image in runs, grey-level non-uniformity, run length non-uniformity, long run emphasis, short run emphasis *(in four directions)*	20

Table 2.5- List of statistical texture features used in this thesis

Run-Length Matrix

The run-length matrix is based on the number of consecutive pixels in a given direction having the same grey-level intensity value [89]. A matrix is composed using the number of intensity levels in the image and the number of pixel runs (i.e. how many consecutive pixels have the same pixel intensity- see Figure 2.17) [121]. This is highly dependent on the number of grey levels in the image and these will often be reduced to increase the counting statistics, at the expense of reduction in textural information in the image [86].

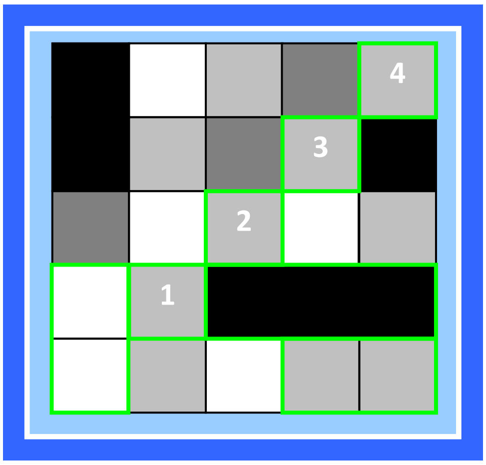


Figure 2.17- Example image showing run length of 4 in 45° direction for light grey pixel intensity

Run length matrices are usually calculated in four directions (horizontal, vertical and along the two diagonals) as shown in Figure 2.17 (see Table 2.5) [124].

2.2.3 Model Based Methods for Texture Analysis

Auto-regressive Model

Model based texture analysis methods assume that the image texture can be represented by a computational model to which parameters can be fitted [86, 89]. The method employed in the MaZda texture analysis software program, as used in this thesis, is the auto-regressive model (ARM) [125, 126]. This calculates grey levels using weighted sums of neighbouring pixel intensities and the auto-regressive parameters are those weights. These are used to establish a relationship and are therefore measures of the statistical similarities between a pixel and its neighbours, which can be related to texture [86, 89]. The weights are denoted by theta (θ_A), while the noise in the image is denoted by sigma, σ_A (see Table 2.6).

			Number of Features calculated
Auto Regressive Model	ARM	theta (θ_A) 1-4, sigma (σ_A)	5

Table 2.6- Auto-regressive model features used in this thesis

2.2.4 Transform Methods for Texture Analysis

Wavelet Transform

Transform methods consider the data in a different space and the method implemented in MaZda, the wavelet transform, analyses the image in frequency space.

In an image, fine detail is represented by high frequency signals while low frequency signals encode the contrast information. How we perceive an image is also dependent on the size at which we view it- i.e. if we view an image at a large scale we notice gross features while smaller details are noted at small scales [117]. This accounts for the dependence of texture perception on image resolution as well as the overall size of the image. The wavelet transform utilises the frequency information and viewing scale to analyse texture content of an image.

			Number of Features calculated
Wavelet Transform	WAV	sub-bands LL, LH, HL, HH at different energies *(up to five energies calculated)*	20

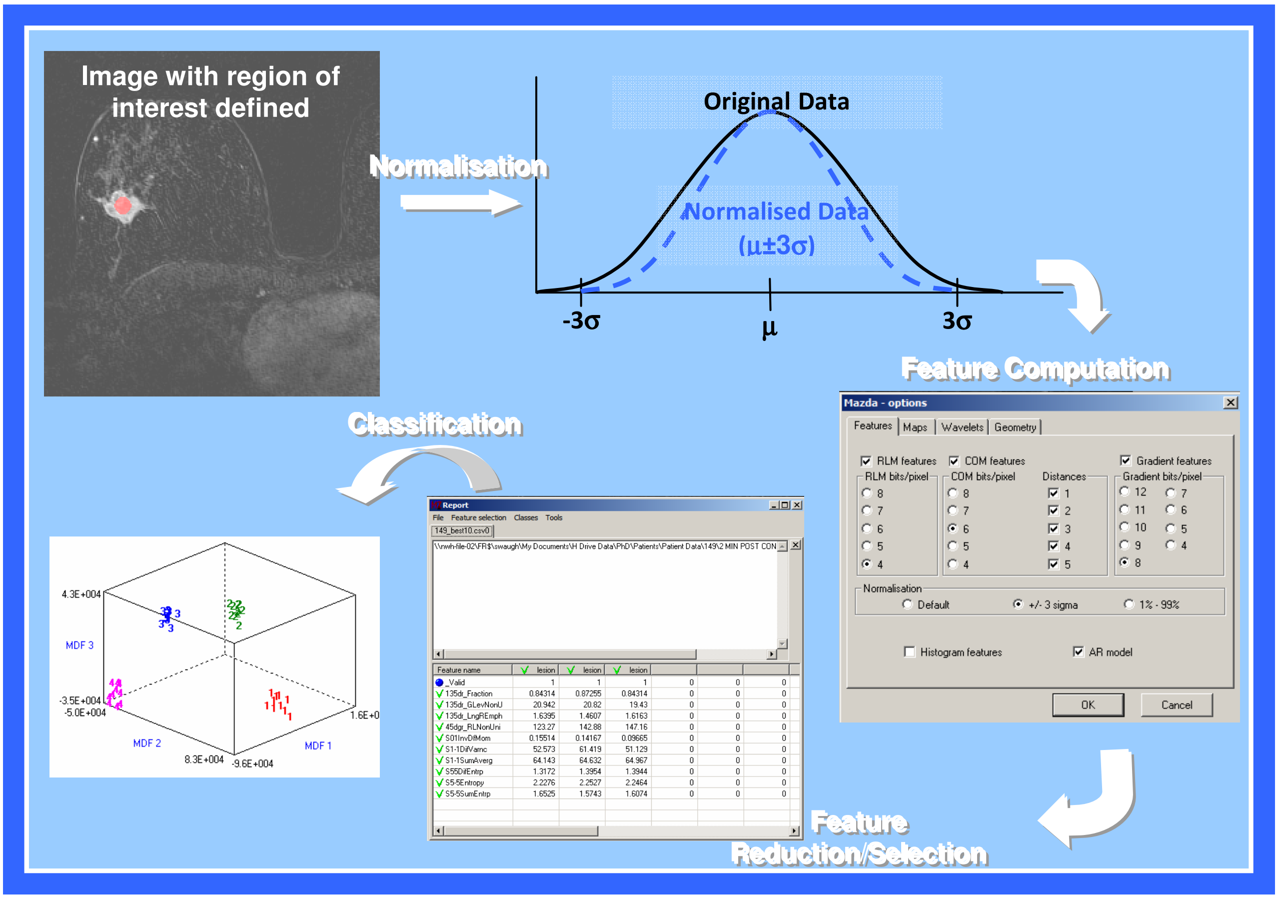
Table 2.7- Wavelet parameters as calculated by MaZda, used in this thesis.

The image is passed through low and high pass filters before being rescaled and again filtered. This provides information on the frequency content of the image at various viewing scales (as shown in Table 2.7). The wavelet energy is then calculated using the signal in each frequency channel, which is directly linked to the frequency content of the whole image and therefore the image texture [86]. The Haar wavelet is the most commonly implemented transform as it allows for the precise localisation of texture, whilst minimising the signal distortion. Wavelet energy is calculated for each scale.

2.3 COMPUTER SOFTWARE USED IN THIS THESIS

2.3.1 Computer Aided Texture analysis

Texture analysis is a computationally demanding technique and there are a number of steps performed in order to obtain meaningful results from the technique (Figure 2.18). Texture analysis within the scope of this thesis has been carried out using MaZda (Technical University of Lodz, Poland) [125-127], which is a well-known and established software program, which is freely available. Additionally, in Chapter Five a custom, independently-written piece of software was used to compare a max-min texture analysis technique with results obtained from MaZda. The computer coding for this can be found in Appendix B.



☑ RLM features
RLM bits/pixel: ○ 8 ○ 7 ○ 6 ○ 5 ● 4

☑ COM features
COM bits/pixel: ○ 8 ○ 7 ● 6 ○ 5 ○ 4
Distances: ☑ 1 ☑ 2 ☑ 3 ☑ 4 ☑ 5

☑ Gradient features
Gradient bits/pixel: ○ 12 ○ 7 ○ 11 ○ 6 ○ 10 ○ 5 ○ 9 ○ 4 ● 8

Normalisation: ○ Default ● +/- 3 sigma ○ 1% - 99%

☐ Histogram features
☑ AR model

\\nwh-file-02\FR$\swaugh\My Documents\H Drive Data\PhD\Patients\Patient Data\149\2 MIN POST CON

Feature name	✓ lesion	✓ lesion	✓ lesion			
● _Valid	1	1	1	0	0	0
✓ 135dr_Fraction	0.84314	0.87255	0.84314	0	0	0
✓ 135dr_GLevNonU	20.942	20.82	19.43	0	0	0
✓ 135dr_LngREmph	1.6395	1.4607	1.6163	0	0	0
✓ 45dgr_RLNonUni	123.27	142.88	147.16	0	0	0
✓ S01InvDfMom	0.15514	0.14167	0.09665	0	0	0
✓ S1-1DifVarnc	52.573	61.419	51.129	0	0	0
✓ S1-1SumAverg	64.143	64.632	64.967	0	0	0
✓ S55DifEntrp	1.3172	1.3954	1.3944	0	0	0
✓ S5-5Entropy	2.2276	2.2527	2.2464	0	0	0
✓ S5-5SumEntrp	1.6525	1.5743	1.6074	0	0	0

Figure 2.18- Graphical representation of the process required for texture analysis and feature classification

Image Normalisation

As there are typically 256 or more grey level intensities contained within an image, it is normal to compress the data into a smaller intensity range in order to simplify further computation of texture features. Performing this step increases the statistical power; however some texture information may be lost within this step. The most common way of performing this step is to rescale the histogram to fit into ±3 standard deviations (σ) of the histogram mean (μ).

Feature Computation

MaZda software calculates features based on the image histogram, co-occurrence matrix, absolute gradient, run-length matrix, auto-regressive model and wavelet transform.

Feature Selection

Due to the large number of features generated by MaZda, all features or only a subset can be selected to export for further analysis.

While it is also possible to perform feature reduction to decrease the dimensionality of the data [118] using various methods such as mutual information, probability of error and the Fisher combination, as these techniques were not used within this work, they have not been defined further.

2.3.2 Data Classification

As texture analysis results in a large number of features being generated to describe regions of interest within an image, classification regimes are employed to look for patterns within the data. The software used in this thesis for classification of features were the packages B11 (Technical University of Lodz, Poland) [128, 129] and Weka (The University of Waikato; Hamilton, New Zealand) [130].

Classification has been performed throughout using the k-nearest neighbour technique (k-NN) where distance functions are calculated to ascribe feature vectors to the class with a minimum distance in feature space. The number of neighbours required in order to ascertain whether a data vector should be assigned to the class (the ‘k’) can be selected according to requirements. Using a higher number of nearest neighbours generally results in a slight reduction in the number of correctly classified vectors, but an increase in the classification certainty.

Validation methods are employed in classification packages to minimise the possibility of classification due to chance, and can be categorised as either internal or external methods. The internal method utilised in this thesis is the cross-validation regime using a ten fold regime. The data is split into a number of ‘folds’, in this case ten. The model is built on $^{9}/_{10}$ of the data and tested on the remaining tenth and this procedure is repeated ten times. This ensures that the best model is built to fit the data supplied. External validation regimes test the model on ‘unseen’ data, and the method employed by Weka is the holdout method. The percentage of data to be training data is selected (this was 66% of data except where explicitly stated to be different) and the model is formulated on this prior to testing on the remainder of the data. This provides a more accurate representation of how well a model can be expected to perform on new data.

Assessment of how well the classification has been performed can be made using the classification accuracy, which normally describes a percentage of the data that has been correctly classified and area under receiver operator curves (ROC). ROC curves are particularly useful as they demonstrate the trade-off between correct identification of positives vectors as a function of misclassifying negatives as positives. They are commonly used in medical diagnostics to determine cut-off values in clinical tests.

2.3.3 Statistical Analysis

As well as considering the individual feature classification accuracies, assessment was also made of the raw feature values themselves. All statistical assessment was performed using SPSS version 21.0 for Windows (IBM Corporation; Armonk, New York) [131].

As data could not be assumed to belong to a normal distribution, Mann Whitney U tests were used for all pair-wise comparisons, and Wilcoxon tests employed in Chapter 7 for paired comparisons. Where comparisons were required between multiple groups, Kruskal Wallis tests were employed.

In all cases, a significant level of $p<0.05$ was utilised, except where explicitly stated to be otherwise.

CHAPTER THREE: PHANTOM VALIDATION STUDY

OUTLINE OF CHAPTER: This chapter deals with the stability of texture measurements by designing and making a standard 'texture phantom'. Assessment is then made of the impact that different MRI scanners and different imaging parameters have on texture analysis outcome.

3.1 INTRODUCTION

One of the main concerns surrounding the use of texture analysis (TA) in routine clinical practice is reports in the published literature indicating that the technique is highly dependent on MRI hardware and acquisition conditions [111, 132-135]. This would suggest that while texture analysis can be used reliably in one centre for image analysis, setups in other centres may result in completely different results.

The main reason cited within the literature for this inconsistency is due to signal to noise ratio (SNR) differences [136, 137]. The aim of this study was therefore to identify first of all whether a routine clinical breast sequence could be used for accurate texture classification, using a custom-designed texture phantom, or whether sequence optimisation was required prior to further studies. The robustness of texture analysis was also investigated by altering acquisition conditions and examining the effect on the accuracy of texture analysis.

3.2 MATERIALS AND METHODS

3.2.1 Breast Phantom

There have been a wide variety of texture phantoms described within the literature, including polystyrene spheres, fine glass tubes, reticulated foam and glass spheres embedded in agarose gel or doped water [111, 132, 133, 138, 139]. On imaging such phantoms, however, macroscopic patterns are often visually apparent and therefore visual classification is possible without the need for a computer-based technique. As the aim of this work was to develop TA as a clinical tool, a phantom was designed that would incorporate texture objects that appeared visually identical on imaging. This would allow assessment of the successfulness of TA between different acquisition conditions, without any bias introduced according to the visual appearance of the images.

Reticulated foam has been used successfully as texture objects in other studies [111, 134, 136] and does not result in any obvious visual differences on resulting images acquired at clinical MRI resolutions. It was therefore felt to be the most appropriate material for composing a phantom for our validation studies.

Foam with specific porosities of 30, 45, 75 and 90 pores per inch (ppi) was sourced to use in the test phantom (Foam Engineers Ltd.; Buckinghamshire, UK)- Figure 3.19 (a-d). The foam samples were submerged in tubes filled with a 2% agarose solution (Sigma Aldrich; St. Louis, MO) held at 70°C. The agarose was doped with 0.2% Magnevist (Bayer Healthcare; Germany) to shorten T_1 relaxation times to values comparable with those measured *in vivo* [140]. As air bubbles can adversely affect the TA outcome by artificially introducing further texture [133] and causing susceptibility artefacts, each piece of foam underwent repeated cycles of compression in the liquid agarose solution followed by a period of settling. Once there was no visual evidence of the presence of air bubbles, the texture objects were cooled slowly in a water bath- Figure 3.19.

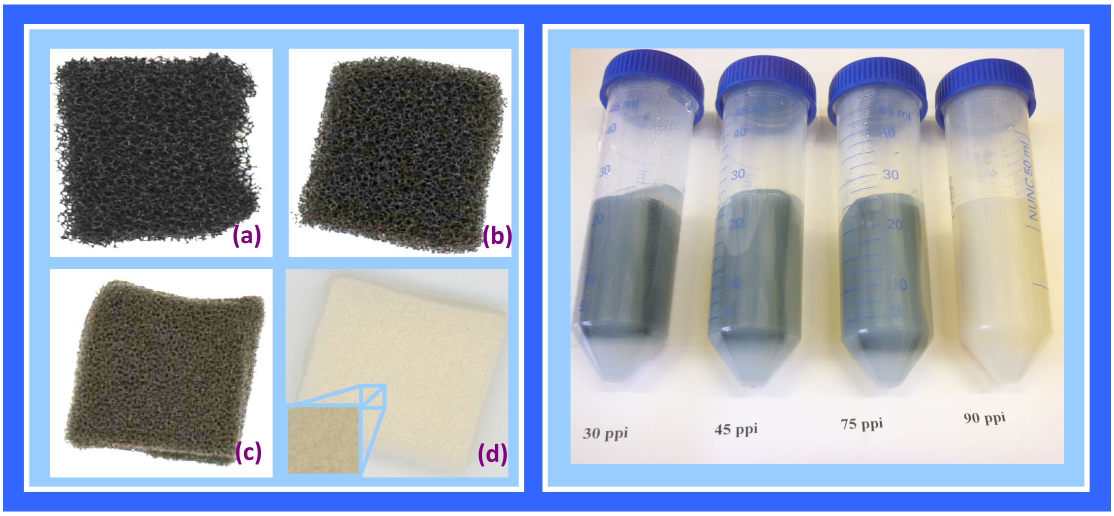


Figure 3.19- Reticulated foam texture phantoms showing different grades of foam ((a)- 30ppi; (b)- 45ppi; (c)- 75ppi; (d)- 90ppi) and then embedded in agarose gel (right hand image)

Once the agarose had set, the phantoms were imaged on a 3.0T Siemens Trio (Siemens; Erlangen, Germany) MRI scanner using a high resolution T_1 FLASH sequence with an in-plane resolution of $0.4 \times 0.4 \text{mm}^2$ to ensure that there were no remaining air bubbles in the gel test objects.

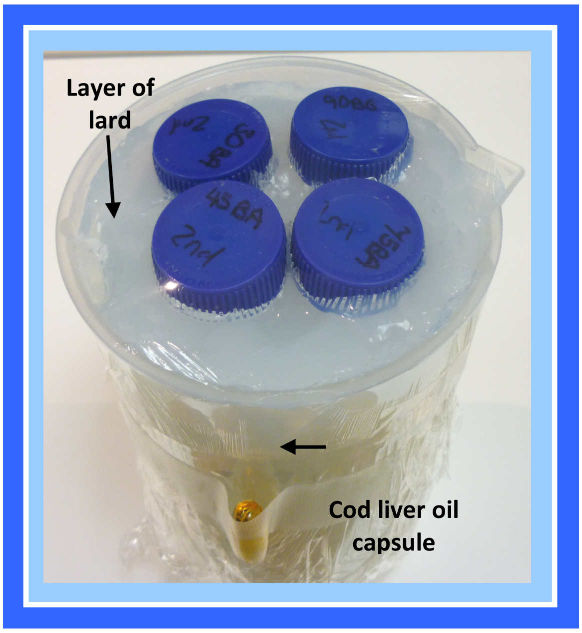


Figure 3.20- Final phantom with four texture objects embedded in the agarose gel

The breast phantom was then made by forming a 5mm layer of cooking lard around a cylindrical container to simulate mammary fat. Agarose gel with the same composition as in the test objects was used to fill the container to mimic fibroglandular tissue, and the test objects were embedded into it (see Figure 3.20). A cod liver oil capsule was placed on the side of the container to ensure reproducible slice positioning when imaging the phantom.

Both T_1 and T_2 values of the agarose and lard were measured on a 1.5 T Siemens Avanto MRI system using standard techniques and compared with clinically obtained values for fibroglandular tissue and mammary fat- shown in Table 3.8.

		T_1 (ms)	*T_2 (ms)*
Phantom Measurements	Agarose Gel	1370	65
	Lard	272	49
Previously reported *in vivo* measurements [140]	Fibroglandular tissue	1266	57
	Mammary fat	296	53

Table 3.8- T_1 and T_2 measurements obtained from phantom, compared with those obtained *in vivo*

Images acquired with the phantom in the 2-channel breast matrix coil and using a standard T2 TSE and T1 FLASH sequence from the routine breast protocol are shown Figure 3.21.

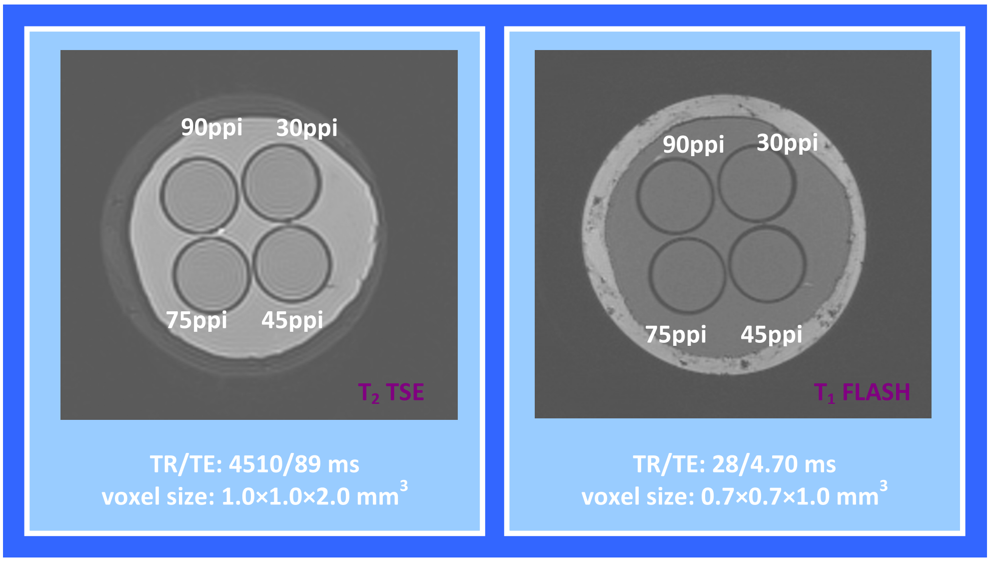


Figure 3.21- Phantom images acquired using typical clinical breast MRI sequences

3.2.2 MR Imaging

All imaging was carried out using a 3D spoiled gradient echo sequence (FLASH), as described in Section 2.1.6. Clinically, this sequence is used as a dynamic acquisition, imaging both breasts axially prior to, during and for a period of time after the administration of a contrast agent to identify regions of increased vascularity, often associated with malignancy. This sequence was chosen for all analysis to be performed on as it is often used to identify lesions in the breast, is incorporated into all breast MR clinical protocols and also because efforts have been made by the European Society of Breast Imaging to standardise its usage across equipment and sites [141].

The standard 3D FLASH dynamic acquisition run clinically acquires a number of volumes sequentially to provide a temporal representation of the contrast uptake. The total scan time is determined by the number of these volumes that are acquired and therefore to minimise imaging time using the phantom, only one volume was acquired.

Figure 3.22- Phantom setup for sequence testing

The phantom was imaged as shown in Figure 3.22, using three different protocols on two different scanners, with acquisition conditions as shown in Table 3.9. A standard quality assurance phantom was used on the contralateral side of the coil to the phantom, to ensure adequate coil loading.

		Protocol 1	*Protocol 2*	***Protocol 3***
Scanner		16-channel Siemens Avanto	32-channel Siemens Trio	32-channel Siemens Trio
Field strength	(T)	1.5	3.0	3.0
Imaging coil		2-channel	7-channel	7-channel
Repetition time	(ms)	3.8	3.8	3.5
Echo time	(ms)	1.24	1.28	1.19
Field of view	(mm)	320×320	320×320	320×320
Matrix		384×384	384×384	256×256
Slice thickness	(mm)	0.83	0.83	2.0
Voxel size	(mm^3)	0.83×0.83×0.83	0.83×0.83×0.83	1.25×1.25×2.0
Flip angle	(°)	6	6	10
Bandwidth	(Hz/px)	650	650	560
Grappa factor		×2	×2	×2
Number of slices		192	192	72
Imaging Time	(s)	61	60	24

Table 3.9- Sequence parameters used for each of the three imaging protocols

Protocol 1 was acquired on a 16-channel 1.5 T Siemens Avanto MRI scanner using a 2-channel breast matrix coil and Protocols 2 and 3 were acquired using a 32-channel 3.0 T Siemens Trio MRI scanner with a 7-element open breast biopsy coil. Both Protocol 1 (P1) and Protocol 2 (P2) were high spatial resolution techniques, with lower temporal resolution, which are used at our institution as routine clinical protocols allowing morphological assessment as well as semi-quantitative enhancement analysis. Protocol 3 (P3) uses a higher temporal resolution with lower spatial resolution which could potentially be used for pharmacokinetic modelling [142-144].

The effect of sequence parameter changes on TA outcome was investigated by altering three of the acquisition parameters most likely to be altered in a clinical setting- the repetition time (TR), the bandwidth/echo time (BW) and the flip angle (α). The echo time (TE) is set as the minimum value for a given bandwidth to minimise acquisition times, as per the manufacturers recommended protocol.

For each protocol the phantom was imaged using the baseline protocols as outlined in Table 3.9, and then four parameter changes were implemented to cover the range likely to be encountered clinically (see Table 3.10). Only one parameter was changed at any given time, while the other parameters remained at their respective baseline values (TR1, BW1, α1).

		P1	*P2*	***P3***
(ms)	TR1	3.80	3.80	3.50
	TR2	4.00	4.00	3.25
	TR3	4.25	4.25	3.75
	TR4	4.50	4.50	4.00
	TR5	4.75	4.75	4.25
(Hz/px/ms)	BW1/ TE	650/1.24	650/1.28	560/1.19
	BW2/ TE	590/1.26	590/1.33	650/1.11
	BW3/ TE	540/1.29	540/1.55	530/1.22
	BW4/ TE	500/1.31	500/1.57	510/1.23
	BW5/ TE	470/1.34	470/1.60	470/1.29
(°)	α1	6	6	10
	α2	5	5	8
	α3	4	4	12
	α4	3	3	14
	α5	2	2	16

Table 3.10- Sequence parameter changes for each of the considered acquisition parameters at each imaging protocol

All images were stored in Siemens DICOM format (.IMA) and were re-saved as true DICOM (.dcm) files for texture analysis by using a filename extension changer.

3.2.3 Texture Analysis

Texture analysis was performed using MaZda version 4.7 [126, 127, 145] and data classification performed using Weka, version 3.6.9 [130].

For texture analysis, circular regions of interest (ROI) were drawn manually with a fixed area of no less than 300 pixels (corresponding to a physical area of approximately 216 mm^2) for P1 and P2, and no less than 170 pixels (corresponding to a physical area of approximately 276 mm^2) for P3. Slight differences in ROI sizes were mainly due to the marginally larger Gibbs artefact (bright and dark lines next to the border of abrupt signal intensity changes) in the images acquired with P3- ascribed to the larger pixel size. These ROIs were placed in the centre of each foam phantom across the ten central slices.

Image data was re-binned according to the model in question using previous experience, reports in the literature [105] and recommendations from the software developers [146] (co-occurrence matrix: 6 bits/pixel, gradient features: 8 bits/pixel, run-length matrix: 4 bits/pixel, wavelet transform: 12 bits/pixel). Grey level normalisation, which is known to minimise the effect of contrast and brightness variations on the outcome of TA [127, 137] was carried out within MaZda by rescaling the histogram to fit within $\mu \pm 3\sigma$ (where μ- grey level mean, σ- grey level standard deviation).

Texture features were calculated using five different techniques: the auto-regressive model (ARM), co-occurrence matrix (COM), absolute gradient (GRA), run length matrix (RLM) and wavelet transform (WAV), as described in Chapter 2.2. The texture features associated with each category are shown in Table 3.11.

			Number of features
Auto-regressive model	ARM	theta (θ) 1-4, sigma (Σ)	5
Co-occurrence matrix	COM	angular second moment, contrast, correlation, difference entropy, difference variance, entropy, inverse difference moment, sum average, sum entropy, sum of squares, sum variance	11 (in four directions and up to interpixel distances of 5)
Absolute gradient	GRA	gradient mean, kurtosis, non-zero, skewness, variance	5
Run-length matrix	RLM	Fraction of image in runs, grey-level non-uniformity, run-length non-uniformity, long run emphasis, short run emphasis	5 (in four directions)
Wavelet transform	WAV	sub-bands LL, LH, HL, HH at five energies	4 (up to five energies)

Table 3.11- Texture features derived from each of the categories used in this study

As there are a total of 220 features that can be calculated using the COM model, these were standardised to limit the number of features calculated. Only texture features for pixel pairs at an interpixel distance of n=1 and in arbitrarily chosen directions of 0° and 45° were calculated, which then gave an input of 22 features from the COM model for classification (two entire datasets of eleven COM features). The interpixel distance was selected as it was believed to be the best for representing fine texture due to the large number of pixel pairs it would produce.

For classification using Weka, cross-validation was utilised in order to increase certainty in the classification outcome. The classification itself was performed using the k-nearest neighbour technique, in which data vectors are classified based on the surrounding vectors in feature space. For this investigation the k=3 nearest neighbour technique was used in order to increase the specificity of the classification. As the true class of the data vector in question was known, the percentage of misclassified vectors was calculated to provide an indication of the successfulness of texture analysis [147].

The aim of this work was two-fold- firstly to identify if four grades of foam within one image could be differentiated using TA for each of the protocols outlined in Table 3.9; and secondly, to determine if changes to baseline acquisition parameters resulted in a measurable difference in the image texture. The rationale for the latter part was to assess whether routine, clinical sequences could be used for TA or whether optimisation was required.

3.2.4 Signal to Noise

To identify whether any differences in TA in response to changing acquisition parameters were real or due to changes in noise levels within the image, signal to noise (SNR) was measured for each sequence parameter change using each protocol.

Measurements were made of the mean signal intensity in each foam sample and the standard deviation of the background noise using circular ROIs and SNR calculated using Equation 3.7.

$$\text{SNR} = 0.655 \times \frac{\text{mean signal}}{\text{std dev background noise}}$$

Equation 3.7

The 0.655 scaling factor is required to correct for the Rayleigh distribution of the noise in magnitude images.

The SNR was plotted as a function of the percentage of misclassified vectors to determine if any relationship could be established.

3.3 RESULTS

Each dataset consisted of 40 datapoints (four different foam porosities over 10 imaging slices) and a total of 13 datasets (baseline parameters plus four of each parameter change) were analysed for each of the five texture models and for each protocol. This led to a total analysis workload of 195 datasets.

An example phantom image is shown in Figure 3.23, which was acquired using P1 with the standard baseline parameters. It is clear from this figure that each foam sample cannot be discriminated visually (Multi-Modality WorkPlace (MMWP) workstation resolution: 1280×1024).

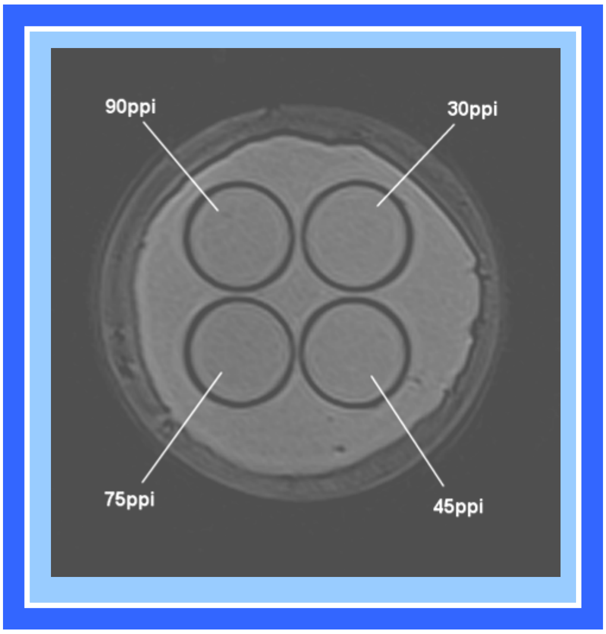


Figure 3.23- MR image acquired using the standard baseline sequence (P1) of the breast phantom composed of four different grades of reticulated foam, as shown

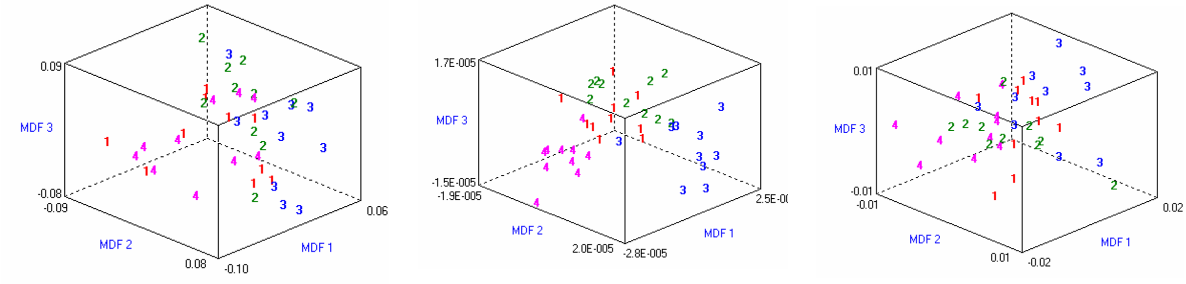
First order statistics were calculated and confirm that the foams could not be differentiated without the use of second order statistics (Mean: 143.63 [141.40-147.16]; Variance: 35.53 [28.68-39.78], Skewness: 0.20 [0.15-0.34]; Kurtosis: 0.28 [0.20-0.42]).

3.3.1 Differentiation of Foam at Each Baseline Imaging Protocol

The ability of TA to discriminate between the four different grades of foam was assessed by considering the percentage of misclassified vectors for each of the individual models. Results are presented in Table 3.12 for the images acquired with baseline parameters for each protocol. Figure 3.24 shows these results graphically in 3-D feature space. It is clear from these results that the wavelet transform, WAV, and the co-occurrence matrix, COM, features performed the best for discriminating between the four foams across all protocols. The run-length matrix, RLM, gradient features, GRA, and auto-regressive model, ARM, resulted in the poorest overall classification.

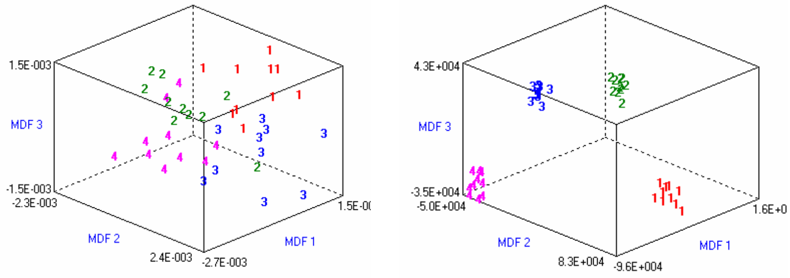
	ARM	*COM*	*GRA*	*RLM*	***WAV***
P1 baseline	52.5	52.5	62.5	77.5	2.5
P2 baseline	22.5	20.0	30.0	17.5	0.0
P3 baseline	27.5	17.5	62.5	62.5	0.0

Table 3.12- Percentage of misclassified vectors for baseline sequence parameters for each protocol and texture analysis model



ARM ***COM*** ***GRA***

1- 90ppi, 2- 75ppi, 3- 30ppi, 4- 45ppi



RLM ***WAV***

Figure 3.24- Graphical representation in 3-D feature space of texture features from each model derived from the baseline images for P2 baseline parameters

3.3.2 Effect of Sequence Parameter Changes

The ability of TA to differentiate between the four foams when using the different sequence parameters outlined in Table 3.10 is presented in Table 3.13. The results are the average percentage rates of

		ARM	*COM*	*GRA*	*RLM*	***WAV***
P1	TR	57.00	67.50	69.00	68.50	2.50
	BW	51.25	50.00	69.00	76.00	1.50
	α	51.50	56.00	61.00	73.50	1.50
P2	TR	21.00	15.00	37.50	28.50	0.00
	BW	21.50	10.00	29.50	23.00	0.00
	α	10.10	12.50	23.50	20.50	0.00
P3	TR	43.95	21.00	64.50	68.50	1.00
	BW	38.00	28.50	54.00	56.50	1.00
	α	32.00	26.50	56.00	62.00	0.50

Table 3.13- Average percentage of misclassified vectors across sequence acquisition parameters for each model at each protocol

misclassification across all parameter changes in each category (i.e. TR1-TR5, BW1-BW5 and α1-α5). Again, WAV features resulted in excellent classification across all sequence parameter changes for each protocol. The COM features also resulted in good classification, particularly for the 3.0 T protocols, and results are presented in Figure 3.25 for each individual sequence parameter change and every protocol for COM features only. While the COM resulted in particularly good classification for P2, which was the 1.5 T parameter matched sequence acquired on the 3.0 T scanner, both the RLM and GRA models appeared to again perform poorly when compared with the other models across all sequence parameters and protocols.

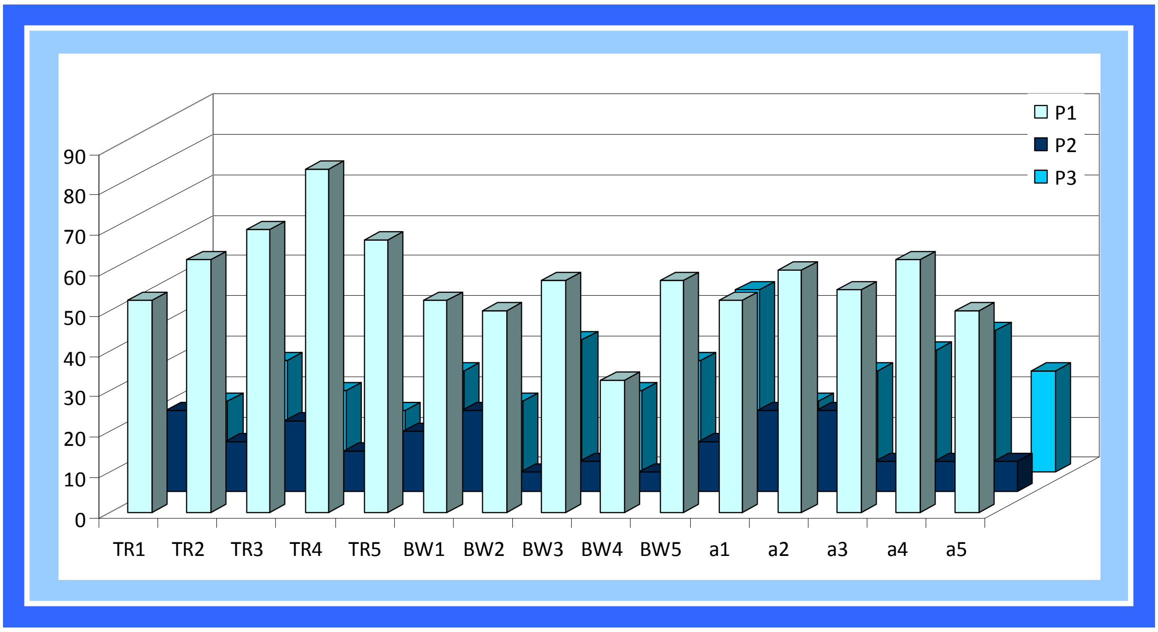


Figure 3.25- Percentage of misclassified vectors for each of the sequence parameters at each protocol for COM features

3.3.3 Comparison of Sequence Parameters

Figure 3.26 shows the comparison of TA when discriminating between two images acquired with different sequence parameters, one of which was baseline. Results are presented graphically only for the COM and WAV as these have been demonstrated in our study to be good at discriminating between the foam phantoms across a range of acquisition conditions.

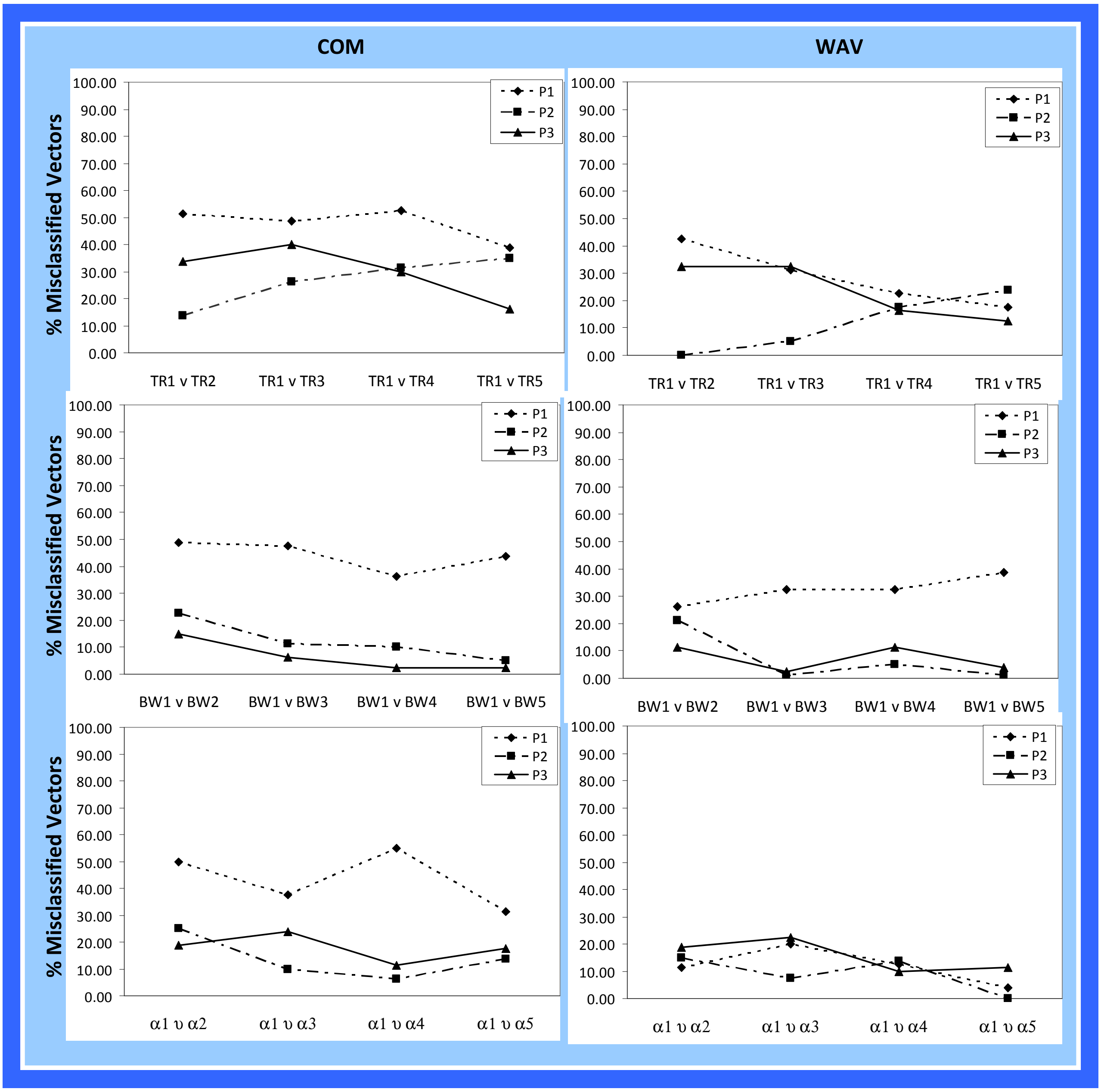


Figure 3.26- Percentage of misclassified vectors calculated when images acquired with different sequence parameters were compared. Data is presented graphically only for the COM and WAV

For variations in the repetition time, P1 resulted in relatively high rates of misclassification, and there was a reduction in the misclassification rates for both P1 and P3 as the TR was changed further from baseline value, suggesting there may be an influence of this parameter change on TA outcome.

When varying the bandwidth/ echo time, P1 resulted in fairly consistent levels of misclassification, while P2 and P3 had a reduction in misclassification as BW/TE was changed further from baseline parameters.

Flip angle changes resulted in a largely unchanged level of misclassification for both COM and WAV features across all changes.

Interpretation of Figure 3.26 shows that the rate of misclassification was highest for both P1 and P3. Protocol 2 consistently resulted in the lowest rates of misclassification and appeared to be most affected by changes to acquisition parameters, suggesting that TA was able to detect differences between two images acquired with different sequence parameters using this protocol.

3.3.4 Effect of SNR on Misclassification

Measurements of SNR are given in Table 3.14 There were significant differences in the measured SNR between protocols P1 vs P2 and P2 vs P3 ($p<0.001$), however there were no significant differences found between P1 and P3. The measured signal to noise ratio was lowest for P2.

Despite differences in the measured SNR, there was no obvious influence on the outcome of TA. For all protocols there was no correlation between the rate of misclassified vectors and the measured SNR ($r^2<0.042$).

	P1	*P2*	***P3***
TR1	49.41	30.85	45.88
TR2	50.10	35.11	61.18
TR3	52.01	41.15	69.94
TR4	63.09	47.78	73.08
TR5	56.16	44.22	84.67
BW1	49.41	30.85	45.88
BW2	53.84	44.67	49.88
BW3	48.07	45.44	57.35
BW4	50.26	51.01	55.83
BW5	50.69	54.14	66.87
α1	49.41	30.85	45.88
α2	58.87	32.58	66.9
α3	45.53	28.10	52.15
α4	47.17	31.90	48.73
α5	39.23	17.51	37.80

Table 3.14- Measured SNR values for all protocols and all sequence parameter changes

3.4 DISCUSSION

One of the main concerns surrounding the use of texture analysis in routine clinical practice is reports within the published literature highlighting the dependence of the technique on MRI hardware and acquisition conditions [111, 132-135].

In order to assess the effect of changing the sequence parameters and further investigate the relationship between SNR and texture analysis outcome, a stable breast-mimicking phantom with texture objects embedded was constructed.

Agarose gel and lard were used to represent the breast and four different grades of foam were used as texture objects.

The aim of the first part of the experiment was to demonstrate whether routine imaging practices could be used to differentiate between the different grades of foam using texture analysis techniques. The findings from the investigation show that TA is able to differentiate between the four grades of foam porosities using standard clinical imaging protocols and imaging techniques on two MRI systems of different field strengths, despite the foam phantoms appearing visually identical on the acquired MR images. Whilst the WAV consistently resulted in accurate classification of the four foams, the COM also resulted in good classification, particularly when considering P2- the high spatial resolution sequence acquired on the 3.0 T scanner. Both the RLM and the GRA features performed poorly for classification of this particular foam phantom within the imaging protocols tested.

When acquisition parameters were changed for each of the imaging protocols in the second part of this investigation, TA was still able to differentiate between the four phantoms and WAV again performed best with very low misclassification for any of the protocols or sequence parameter changes. Wavelet transforms have proved to be a valuable tool for texture analysis as frequency components are analysed at different ‘scales’, and as texture is critically linked to the scale at which the data is viewed, this is likely to a particularly relevant texture model- especially when no visual differences are observed.

The COM model, which has been reported in the literature to be the most effective method for performing texture analysis on medical images [121], provides highest misclassification for P1, the 1.5 T high spatial resolution protocol. For this case, the repetition time appeared to affect the classification accuracy the most (see Figure 3.26). COM performed much better using the high field protocols (both P2 and P3) compared with the lower field strength protocol, and particularly so for P2, the 1.5T matched high spatial resolution protocol, which showed good classification across all sequence parameter changes. For the high temporal resolution sequence acquired at 3.0 T (P3), the changes to the bandwidth/ echo time appeared to have a more dramatic impact on the classification success. Our results show that the repetition time (TR) appears to have the biggest impact on texture analysis when TA is used to detect texture differences between images acquired with different imaging parameters. This was consistent for both the COM and WAV features using P2. For BW/TE, there was again a similar trend between COM and WAV features for P2, with an increasing classification performance for decreasing BW values relative to the baseline suggesting TA was able to detect differences between the two images.

Protocol 1 (P1) appeared to be the least affected by changes in sequence parameters with the overall highest rates of misclassification across all sequence parameter changes for both COM and WAV texture analysis models.

Measurements of SNR were carried out and no significant correlations were found between the percentages of misclassified vectors and the measured SNR. The lowest SNR was measured for P2, and this is attributed to the flip angle selected, which potentially may result in lower signal at 3.0 T compared with 1.5 T.

The limitations of this study include the use of only three different sequence parameters. These acquisition parameters were chosen as those most likely to be changed clinically and only very small changes to the baseline values were used, as again this is most reflective of standard clinical practice. Whilst bandwidth and echo time could have been changed independently, the manufacturer's recommended protocol suggested that a minimum TE value should be used for a given BW.

This study, whilst not exhaustive, demonstrated unambiguously that TA could be successfully used on a breast-mimicking phantom with a standard clinical setup and more extensive changes in acquisition conditions were therefore not deemed necessary. It should be noted that while a 2° flip angle was considered in this study for completeness, it is unlikely to be of use clinically due to the low resulting SNR. While the foam phantoms are unlikely to directly replicate similar textures exhibited in the breast, this study was designed to look at comparability of texture outcomes acquired under different imaging conditions, rather than to truly mimic the underlying textural patterns likely to be encountered in a human breast. There is, however, potential for future development of such a phantom, should this technique become more widely utilised and also to perform quality assurance measures for multi-centre studies.

When considering TA using the COM, we limited the range of calculated features to just 22 by using an interpixel distance of one pixel in only two arbitrarily chosen directions of 0° and 45°. MaZda has a built-in option for selecting the ‘best features’ to describe the data using methods such as the probability of error and Fisher coefficients [126]. However, in this study we wanted to consider a standard set of features to compare between all analyses, and therefore we did not use this option as it was likely to return different features for each individual analysis. Accordingly, the results are likely to reflect a slightly higher rate of misclassification than could be achieved using a set of automatically selected best features to optimise any given dataset. The directions of 0° and 45° were chosen at random and could have been any other combination of directions. This method of utilising two full sets of COM features for standard comparison between analyses has been previously reported by Mayerhoefer et al [135] and was felt to be most appropriate for this study as well.

An additional limitation in the analysis phase is the use of slightly different ROI sizes between high spatial resolution and lower spatial resolution protocols. This may have an effect on the outcome of TA due to different counting statistics, but was unavoidable due to the presence of the ringing artefact, which could result in artificial texture properties if included within the ROI.

Finally, we acquired data on both 1.5 T and 3.0 T systems; however both of these systems use different types of RF coil: a 2-element breast matrix on the 1.5 T, and a 7-element on the 3.0 T. While this is not ideal to form a comparison of two dissimilar coils, it is frequently the case in clinical practice where two systems are equipped with different imaging setups. As there was no opportunity for a direct comparison of both coils on the same MR imaging system, a potential drawback of this study is the inability to directly assess the effect of the field strength on the outcome of TA. While there were measurable differences in the SNR of the final images from the two imaging setups, these appeared to have no correlation with the outcome of texture analysis classification accuracy. It is possible, however, that there were subtle underlying differences in the images not possible to verify from this study due to different coil designs. Further investigation would therefore be required to consider the precise impact of coil architecture on the outcome of texture analysis.

The data presented supports previous studies that have also investigated the outcome of TA for various acquisition techniques, using spin echo sequences

[133, 135]. Our study was concerned with the use of a fast 3D gradient echo sequence with typical clinical imaging parameters and is in agreement with those of Jirak et al. that spatial resolution appears to be the main factor influencing the success of TA [133]. Field strength also appears to have an influence on TA, with protocols matched between a 1.5 T and a 3.0 T scanner suggesting that TA is more consistent on a 3.0 T scanner. While we would expect the most likely explanation for this finding to be attributed to the higher SNR, our measurements were inconclusive when considering the measured SNR in relation to the rate of misclassified vectors, which makes it difficult to explain objectively why there may be an improvement of TA classification using P2 when compared with P1. However, as previously noted, the RF coil used with the 3.0 T system was a 7-element design, whereas for the 1.5 T system a two-element coil was used. Although there may be no correlation between measured SNR and texture data classification accuracy, it is possible that there are other factors to consider, such as the structure of noise across the final image, which could contribute to the improvement in TA. Coil homogeneity could also have a role in the outcome of TA; however this was not investigated within the scope of this work.

CHAPTER SUMMARY: This chapter has demonstrated that texture analysis can reliably differentiate between four grades of foam, despite visually identical appearance on MR imaging. Both co-occurrence matrix and wavelet features have performed best over a range of acquisition parameter changes, and these changes appear to have relatively little influence on the outcome of TA. There appears to be no obvious correlation between outcome of texture analysis with signal-to-noise ratio.

CHAPTER FOUR: RETROSPECTIVE PATIENT STUDY

OUTLINE OF CHAPTER: The previous chapter dealt with the stability and robustness of texture analysis measurements using standard clinical protocols on different scanners by way of a phantom. This chapter considers the use of texture analysis on patients, including the reproducibility of the technique and the usefulness in detection and classification of disease.

4.1 INTRODUCTION

The variable specificity of breast MRI has been widely reported in the literature [43, 46] due to contrast uptake in healthy, hormonal glandular tissue and the overlap in contrast kinetics between benign and malignant disease. Breast cancer has distinctive growth patterns that are used in morphological and histological classification and with work based on 20 patients, Holli et al. suggested that texture analysis may be able to differentiate between lobular and ductal carcinomas of the breast [110].

Recently there has been a drive towards targeted patient treatments based on the molecular classifications of breast cancers [21] as clinically categorised by estogen (ER), progesterone (PR) and human epidermal growth factor 2 (HER2) receptor status and tumour grade. The main subtypes (Luminal, HER2 overexpressed (HER2) and triple negative (TNBC)) exhibit different growth patterns [22, 23] and require different therapeutic regimens.

As there is an increasing drive towards the use of MRI in multidisciplinary decision making for breast cancer, we sought to investigate the use of texture analysis in the detection and classification of breast cancer. We have considered differentiation between normal and malignant tissue, considered malignant and benign lesions and then investigated the use of TA in classification of disease in various ways, including traditional histological subtyping, receptor status categorisation and finally in accordance with full molecular subtyping.

4.2 MATERIALS AND METHODS

4.2.1 Subjects

All women with a primary, operable breast cancer who were referred for a Breast MRI examination between August 2009 and June 2012 were included in this study. All examinations were performed following diagnostic core biopsy 10-21 days prior to MRI. Patients were referred as part of standard clinical care and not actively recruited and only images from patients who agreed for their examination to be used for research and developmental purposes were included. As there was no additional imaging performed and the results of this research had no implications on individual patient care, ethics approval was not required, however Caldicott approval was sought at a local level and approved (Appendix C).

Only lesions greater than 8mm have been included in the following study to enable a standard region of interest for texture analysis to be performed across all datasets, and any examinations where patients moved were excluded from the dataset.

MRI image analysis was performed blinded to histopathological core biopsy data. Cancer subtypes, grading, receptor status and immunohistochemical (IHC) subtypes were obtained from the histopathological reports. Molecular subtype was assessed based on oestrogen, progesterone and HER2 receptor status, with in situ hybridisation for equivocal HER2.

4.2.2 MR Imaging

MR imaging was performed post-biopsy, which was carried out using ultrasound-guided 14 gauge needles as per routine clinical care. Either a 1.5T (Siemens Avanto) or 3.0T (Siemens Trio) MRI scanner with the set-up as described in Chapter 3.2.2 was used. Patients were imaged in a head-first and prone position (Figure 4.27). After the acquisition of standard T_1 and T_2- weighted sequences, a 3D FLASH volumetric sequence was acquired axially through both breasts with an acquisition time of

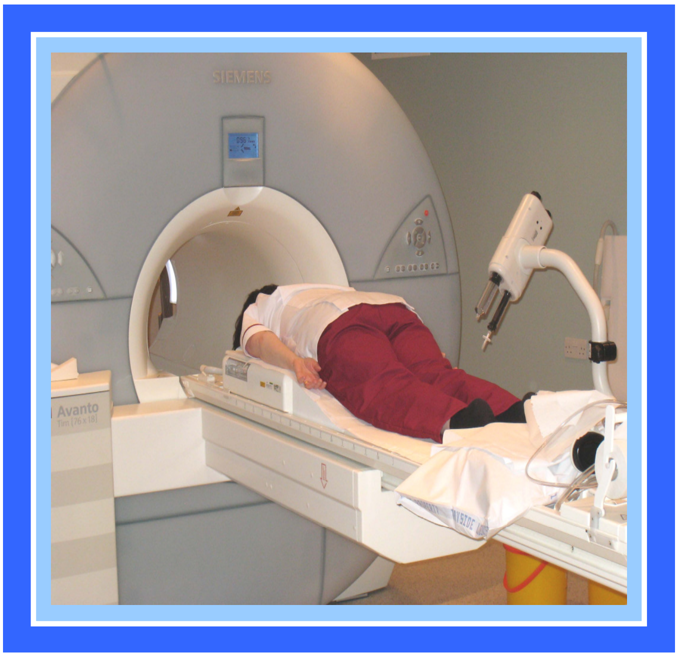


Figure 4.27- Patient positioning for breast MRI examination (setup shown is on 1.5T Siemens Avanto using a 2-channel breast matrix coil)

62s. The complete dynamic acquisition consisted of 8 volumetric acquisitions (total imaging time of 8 minutes 16 seconds) with two volumes acquired prior to the administration of the contrast agent, which was injected using a power injector pump (MedRad; Pittsburgh, PA) (Figure 4.27). All patients received a 0.1mmol/kg dose of Dotarem (Guerbet Laboratories; Villepinte, France) injected at 2.0 ml/s, followed by a 20 ml flush of saline solution at the same injection rate.

Subtracted volumes and maximum intensity projection (MIP) images were generated automatically by the scanner software and DICOM images from the post contrast subtraction acquisitions were exported for analysis.

4.2.3 Texture Analysis and Classification

Texture analysis was performed using MaZda 4.7, with image data re-binned and normalised as previously described (Chapter 3.2.3).

Slices with maximum lesion diameter were selected and square regions of interest (ROI) of 100 pixels (10×10) were drawn manually on sequential slices according to the experiments described. This ROI size was selected as the minimum reported number of pixels required for successful texture analysis [94] to maximise the number of lesions that could be included within this study. Identical ROIs were also drawn in the contralateral, un-enhancing breast.

For all patient work, only the co-occurrence matrix features (COM) were considered in accordance with other works reporting the use of texture analysis in MRI [104, 105]. These were calculated in four directions (0°, 45°, 90° and 135°) and using inter-pixel distances of n=1, 2, 3, 4 and 5, resulting in a total of 220 calculated features. Despite the previous success using wavelets for phantom work, they were found to provide particularly poor results for all patient work and therefore were not considered further.

Classification was performed using Weka, with a k-NN technique (k=3) and a 10-fold internal cross-validation routine. Assessment of classification accuracy was reported in terms of correctly classified data and area under ROC curves. Statistical assessment of raw feature values was also assessed using either Mann-Whitney U tests or Kruskal Wallis tests in SPSS, with $p<0.05$ considered to be significant.

4.2.4 Signal to Noise

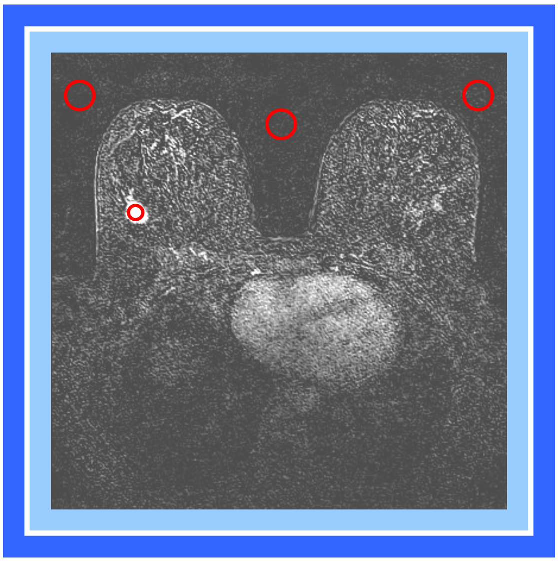


Figure 4.28- ROI positioning for SNR measurements in lesion for each patient.

As the data included in this study was acquired using two different MRI scanners, SNR measurements were performed for all cases to ensure that no significant differences were measured between each classification of cancer type which may have influenced the classification of texture data. In each case, ROIs were placed in the central slice of the lesion and three in background noise in lateral left and right regions as well as medially, between the breasts (Figure 4.28). SNR was then calculated using Equation 3.7.

4.2.5 Aims of Study

Using this retrospective data set, a number of different considerations were carried out as detailed below.

Patient Validation Study

A validation study was first of all performed on 20 randomly selected patients to identify the most appropriate analysis technique. In each case the 1, 2 and 3 minute subtracted contrast volumes were used and ROIs placed in a lesion and normal tissue. Comparisons were made between regular, square ROIs and irregular ROIs, conforming to the shape of the lesion as well as the number of slices used in the analysis- 3, 5 or 10 slices. Mann Whitney U tests were carried out in SPSS to compare statistical differences between lesion and normal tissue, with $p<0.001$ considered significant. Paired t-tests were then used to compare each of the validation criteria to ascertain the optimal analysis criterion ($p<0.05$).

Reproducibility

In order for a clinical tool to be useful and applicable, it must be reproducible for all readers. To assess this, 50 lesions were randomly selected from the databank and analysed twice by one observer (Obs1; SAW) to assess intra-observer reproducibility. An expert MaZda software user, but a novice to breast imaging, was also asked to perform TA on the same 50 lesions to provide a measure of inter-observer reproducibility (Obs2; RAL).

Correlation of the results was then carried out using Bland-Altman plots and also calculating intra-class correlation coefficients using SPSS.

Malignant vs. Normal Tissue

Using the entire dataset of lesions, and analysing in accordance with findings from the validation study, classification and statistical assessment was made as to whether TA could differentiate between normal and malignant tissue and therefore potentially be of use in the detection of breast cancer.

Malignant vs. Benign Lesion

Within the general population, it is estimated that 7-13% of women will have benign lesions such as fibroadenomas [148], which tend to enhance rapidly on breast MR and can mimic pathology hence accounting for up to 50% of breast biopsies performed [148]. Therefore it was expected that in our cohort of patients a number of benign lesions would be identified on MR imaging. Again, classification and statistical assessment was used to identify if TA can differentiate between malignant and benign lesions, and therefore provide a measure of how applicable TA could be in increasing specificity of breast MR for inexperienced readers.

Classifications of Malignancy

As breast cancer is such a heterogeneous group, there are a number of methods for classification of malignancy- either in terms of invasive or in situ, different histological subtypes, grading or full IHC classification.

Pair-wise and global classification, using k-NN; and statistical assessment, using Mann-Whitney U and Kruskal Wallis tests respectively, were performed to assess how well TA could be used in the discrimination of cancer subtypes.

Data was considered as:

- Cancer subtype (ductal; lobular; DCIS)
- Cancer histological grade (grade 1, 2 or 3)
- Immunohistochemical subtype (ER status; HER2 status, Triple Negative (TN) status)
- Molecular subtype of invasive cancers (Luminal; HER2; TNBC)

4.3 RESULTS

4.3.1 Patient Cohort

From our cohort of patients recruited over a 35 month period, there were a total of 148 lesions identified in 200 symptomatic patients (age range 30-81; mean age 55 years), that could be used in our study- i.e. were over 8mm, had full pathological data available obtained within a month of the MRI examination and had consented for images to be used for research purposes. A total of 93 lesions had to be excluded as they did not meet inclusion criteria (i.e. were less than 8mm in size, had incomplete pathological data, did not consent for images to be used or there was gross patient movement on the MRI images). A summary of the lesion characteristics is presented in Table 4.15.

Of these lesions, 108 were identified on examinations that had taken place on the 1.5 T system and 40 were identified on 3.0 T examinations.

Patient Age	30-51 (median 54 years)
Maximum lesion diameter	12.6 mm (range: 8.0-25.2mm)
Histological Subtype	Ductal: 92; Lobular: 45; DCIS: 7
Histological Grade	Ductal- Gd1: 1; Gd2: 30; Gd3: 61
	Lobular- Gd2: 35; Gd3: 10
Molecular Subtype	Luminal: 97; HER2: 26; TNBC: 17

Table 4.15- Summary of population characteristics of lesions included in retrospective set.

4.3.2 Patient Validation

Of the 20 lesions randomly selected from the patient cohort, there were 18 patients with invasive cancer and 2 with in situ cancer.

The number of slices to be used for TA was first of all investigated by considering the number of significantly different features for the COM model between malignant and normal tissue using 3, 5 and 10 slices. Statistical comparison using a student's t-test showed that 10 slices appeared to provide significantly better discrimination compared with only 3 slices (Table 4.16). However, as the analysis regime was required to be identical across all lesions, this indicated that all lesions less than 8.3 mm (10 slices, 0.83mm slice thickness) in the superior-inferior orientation would have to be discarded, further reducing our data set. We therefore concluded that as there were no significant differences between the 3 and 5 slice datasets, that to allow a quick, easily applied regime, 3 slices would be used for texture analysis for this work.

	3 vs. 5 slices	*3 vs. 10 slices*	*5 vs. 10 slices*
p-value	0.107	<0.001	p=0.513

Table 4.16- t-test results from comparisons of texture analysis carried out using 3, 5 and 10 slices. Results show significant differences between number of significant COM parameters

The optimum time after contrast administration was considered by analysing 1-, 2- and 3-minute post contrast subtraction images and regular (square) and

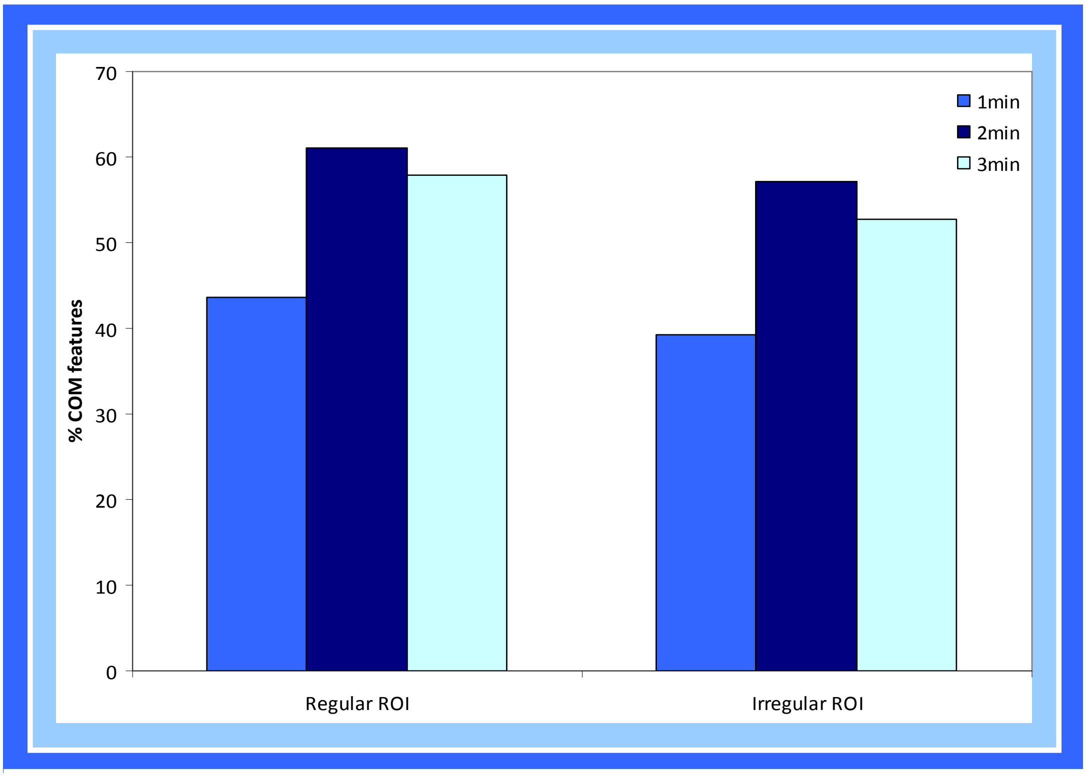


Figure 4.29- Graphical summary showing percentage of significant features between malignant and normal tissue for different post-contrast subtraction images using both regular and irregular ROIs

irregular (conforming to lesion shape) ROIs were drawn on each of the three central slices of the lesions in the validation set. The number of significantly different features between normal and malignant tissue were calculated ($p<0.001$; Mann Whitney U) and are displayed in Figure 4.29 as the percentage of total COM features overall.

From these results, the 2-minute post contrast subtraction scan analysed using a regular ROI (of 10×10 pixels) results in the highest number of significantly different features. There were no significant differences found between the analysis performed at each time-point ($p>0.05$, students t-test), however it should be noted that only 18 data-points were available for the 1-minute post-contrast series due to 2 lesions not enhancing early. There was also no significant difference found between using regular and irregular ROIs ($p=0.053$; students t-test).

The conclusion was therefore that all analysis would be performed using square ROIs of 10×10 pixels on the 2-minute post contrast subtraction series, averaged across the three central slices of the lesion.

4.3.3 Reproducibility

Results show that the overall intra-observer reproducibility obtained by Obs1 demonstrated an excellent intra-class correlation coefficient of 0.870. The intra-class correlation coefficient for entropy features was 0.832 for intra-observer measurements.

Coefficient of variation (CoV) and coefficient of repeatability (CoR) were calculated using Equation 4.8 and Equation 4.9 respectively and are presented for each COM texture feature in Table 4.17.

$$\mathrm{CoV} = \frac{\text{standard deviation}}{\text{average}}$$

Equation 4.8

$$\mathrm{CoR} = 1.96 \times \sqrt{\frac{(\text{measure}_1 - \text{measure}_2)^2}{(n-1)}}$$

Equation 4.9

	Average value	*Average difference*	*CoV (%)*	*CoR*
Angular 2nd Moment	0.028	0.001	80.0	0.046
Contrast	199.827	578.044	10.0	47.123
Correlation	0.086	0.007	123.4	0.164
Diff Entropy	1.125	0.049	18.8	0.436
Diff Variance	65.259	178.406	12.9	26.179
Entropy	1.751	0.120	24.0	0.678
Inverse Diff Moment	0.081	0.0003	17.6	0.034
Sum Average	62.399	42.005	6.8	12.703
Sum Entropy	1.278	0.071	22.4	0.523
Sum of Squares	99.464	157.180	8.9	24.573
Sum Variance	198.028	1183.949	17.2	67.441

Table 4.17- Repeatability parameters for each COM feature as calculated as intra-observer repeatability performed on 50 test cases

Bland-Altman plots were used in order to assess the agreement between the two measurements [149]. An example plot for all the entropy features is shown in Figure 4.30.

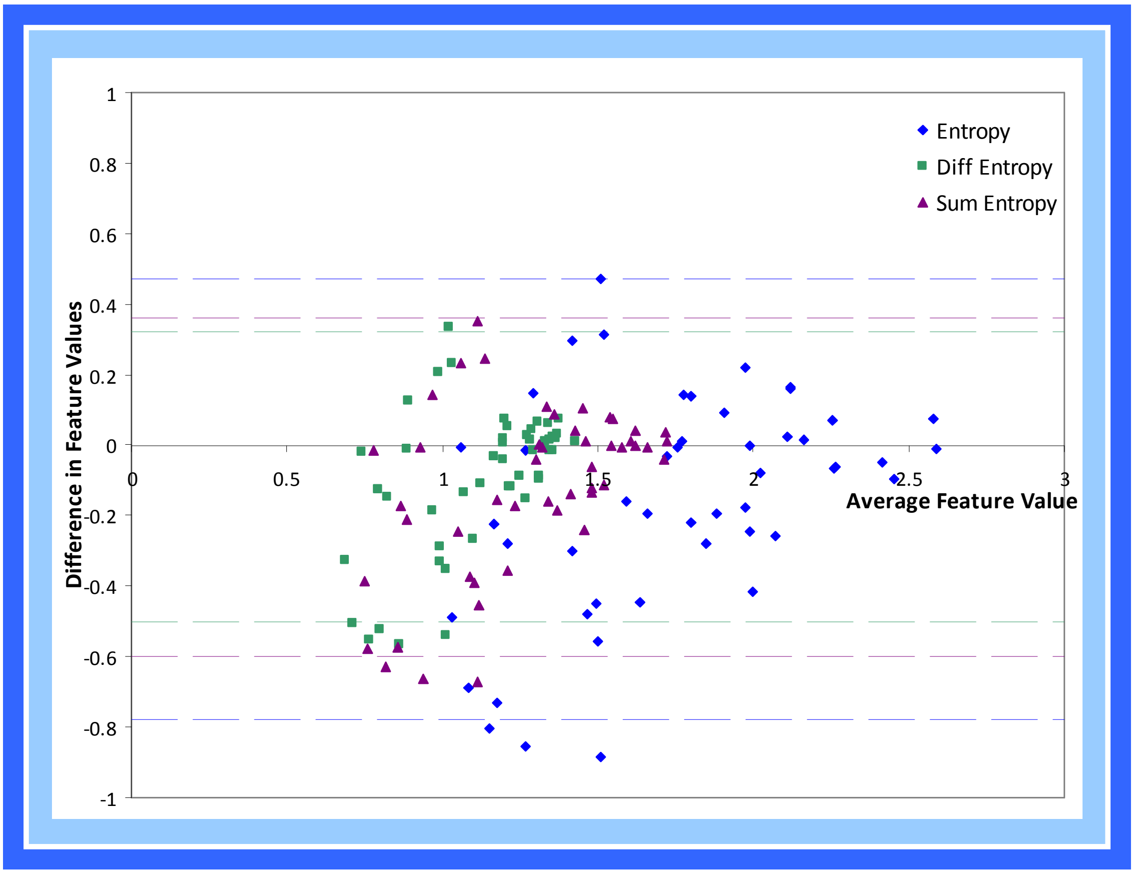


Figure 4.30- Bland Altman plot representing repeatability for intra-observer (Obs1) measurements of entropy features in 50 lesions. Dashed lines indicate ±2 standard deviations of the average feature values

Inter-observer measurements also demonstrated an excellent intra-class correlation of 0.881. For entropy features alone, the intra-class correlation coefficient was 0.834 for inter-observer repeatability measurements. Values for CoV and CoR were calculated according to Equation 4.8 and Equation 4.9 and are presented in Table 4.18.

	Average value	*Average difference*	*CoV (%)*	*CoR*
Angular 2nd Moment	0.017	0.011	73.0	0.066
Contrast	187.699	24.296	12.7	104.479
Correlation	0.064	0.102	137.2	0.349
Diff Entropy	1.221	0.101	10.9	0.581
Diff Variance	62.644	6.203	15.7	35.157
Entropy	1.960	0.264	15.3	1.035
Inverse Diff Moment	0.092	0.017	17.7	0.080
Sum Average	62.825	1.719	10.7	25.794
Sum Entropy	1.422	0.170	13.2	0.767
Sum of Squares	99.624	2.122	11.0	47.494
Sum Variance	210.797	15.806	15.3	128.795

Table 4.18- Repeatability parameters as calculated for comparing results from Obs1 and Obs2 to demonstrate inter-observer repeatability of TA parameters

A Bland Altman plot for Obs1-Obs2 inter-observer repeatability for the entropy features is shown in Figure 4.31. As the values do not appear to be centred around the x-axis, this suggests a small and systematic skew of measures between observers.

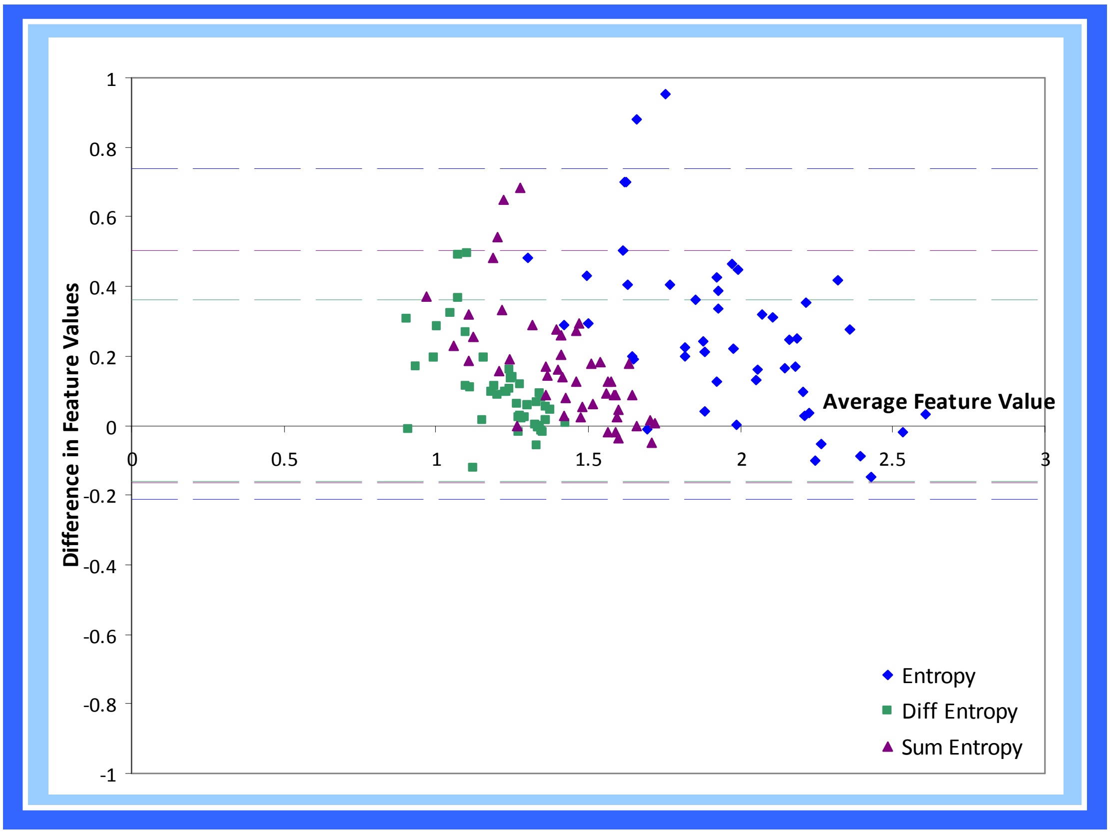


Figure 4.31- Bland Altman plot showing repeatability for inter-observer (Obs1-Obs2) measurements of entropy features. . Dashed lines indicate ±2 standard deviations of the average feature values

4.3.4 Malignant vs. Normal

Comparison was made between normal and malignant tissue. For all COM features as calculated the difference was found to be significant ($p<0.001$; Mann Whitney U). Classification results demonstrated a 97.9% accuracy with ROC= 0.998. When only the entropy features were considered (entropy, sum entropy and difference entropy), the difference was still significant ($p<0.001$; Mann Whitney U) and classification was improved with 99.8% of data correctly classified with an ROC area of 0.998.

These results are demonstrated graphically in Figure 4.32 in a box-whisker plot. The difference between normal and malignant tissue can readily be seen from this plot. Malignant tissue demonstrates statistically, significantly higher values of entropy, sum entropy and difference entropy, suggesting greater heterogeneity in pixel intensity compared with normal tissue.

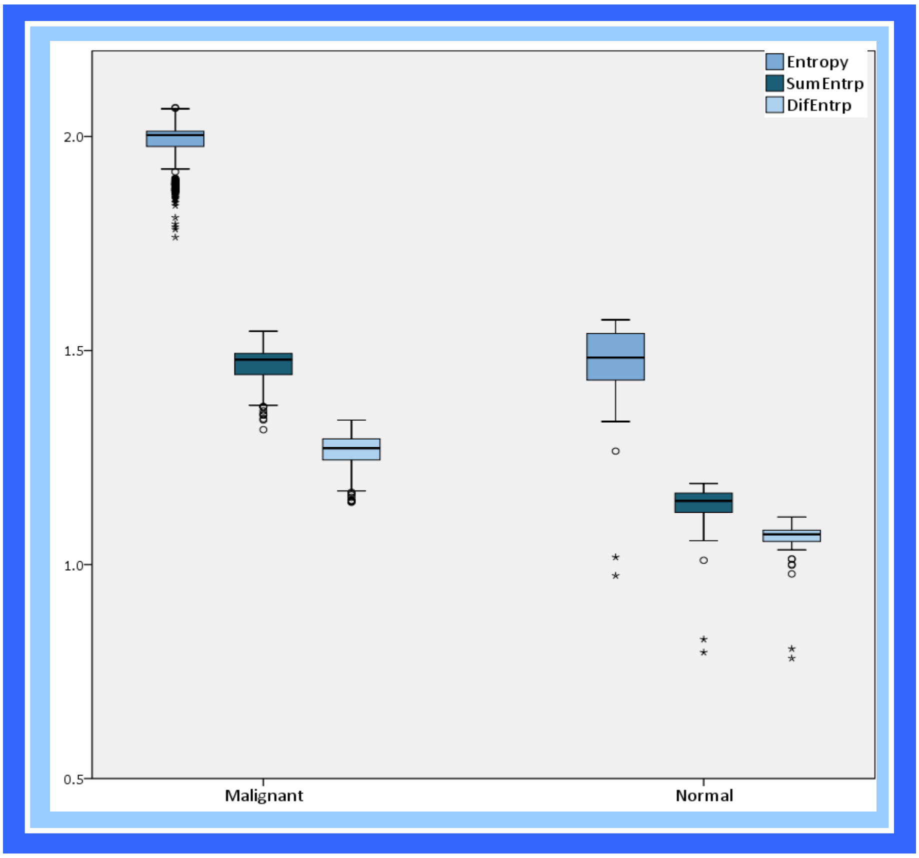


Figure 4.32- Box-Whisker plot showing entropy-based features for both malignant and normal tissue. The differences are statistically significant (p<0.001)

4.3.5 Malignant vs. Benign

In our cohort of 200 patients recruited for this retrospective study, a total of 21 enhancing lesions were identified that were subsequently confirmed to be benign.

These were a mixture of fibroadenomas (16), myxoid fibroadenomas (3), papilloma (1) and one case where the lesion was not visible at follow-up examination.

Figure 4.33 shows two examples of benign lesions identified on MRI examinations. Most of the enhancing benign lesions show similar enhancement patterns to regions of suspicion and particularly in cases where the patient already has a known malignancy (e.g. Figure 4.33(b)), it is imperative to follow these up to determine their nature.

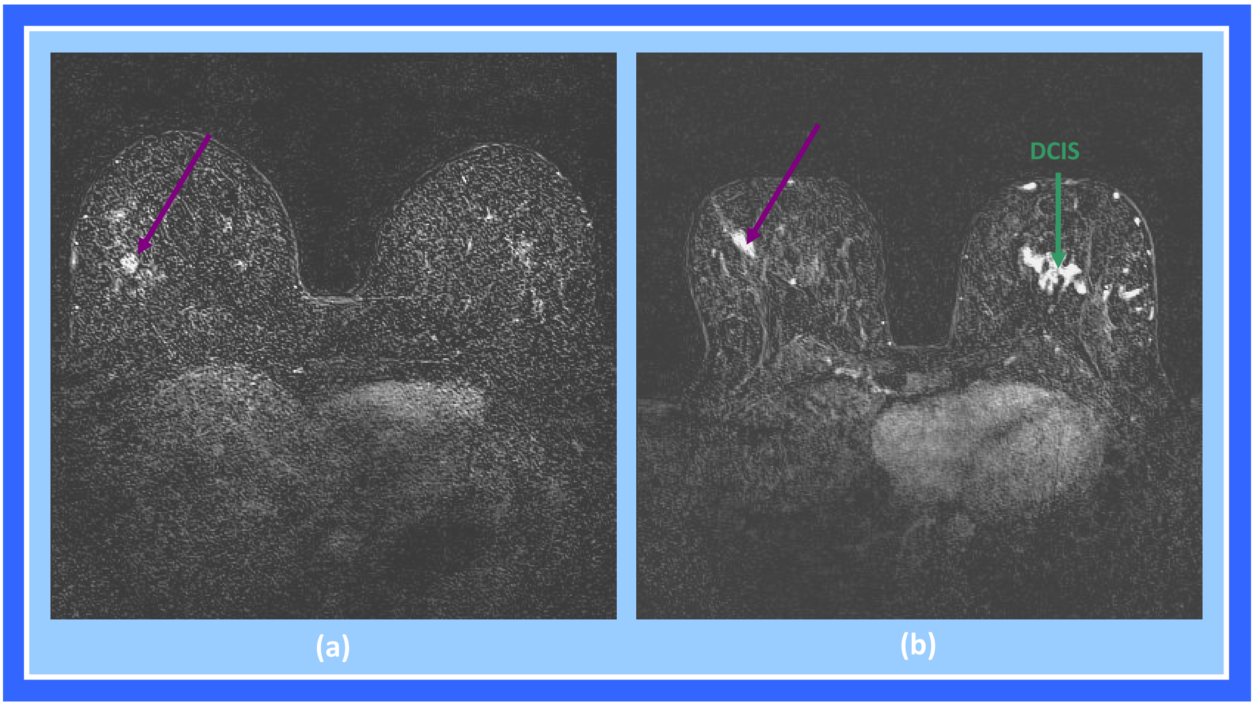


Figure 4.33- Benign lesions identified in the right breast as shown by purple arrows
(a) myxoid fibroadenoma that enhanced and demonstrated washout
(b) patient presenting with extensive high-grade DCIS in the left breast (as shown by green arrow) who also had enhancement in the right breast. This was later confirmed by ultrasound biopsy to be fibroadenomatoid change.

Normal and benign tissue data was added to the tissue from the malignant data obtained in 4.3.3 in order to compare the three types of tissue. Classification results were excellent (see Table 4.19) with 100% classification accuracy in all comparisons and ROC areas of 1.000. Statistical comparisons also resulted in significant differences for all COM features as shown in the last column of Table 4.19.

	Classification Accuracy (%)	*ROC*	*Statistical Testing*
Benign vs Normal	100%	1.000	p<0.001
Benign vs Malignant	100%	1.000	p<0.001
Benign vs Malignant vs Normal	100%	1.000	p<0.001*

Table 4.19- Results of classification and statistical analysis for normal, malignant and benign tissue. Statistical testing was performed using a Mann Whitney U test except for (*) where more than two comparison groups required a Kruskal Wallis test to be employed

Entropy values are represented graphically in Figure 4.34. Entropy features are demonstrated to be lowest for the benign lesions, suggesting a more homogeneous pixel distribution compared with both normal and malignant tissue.

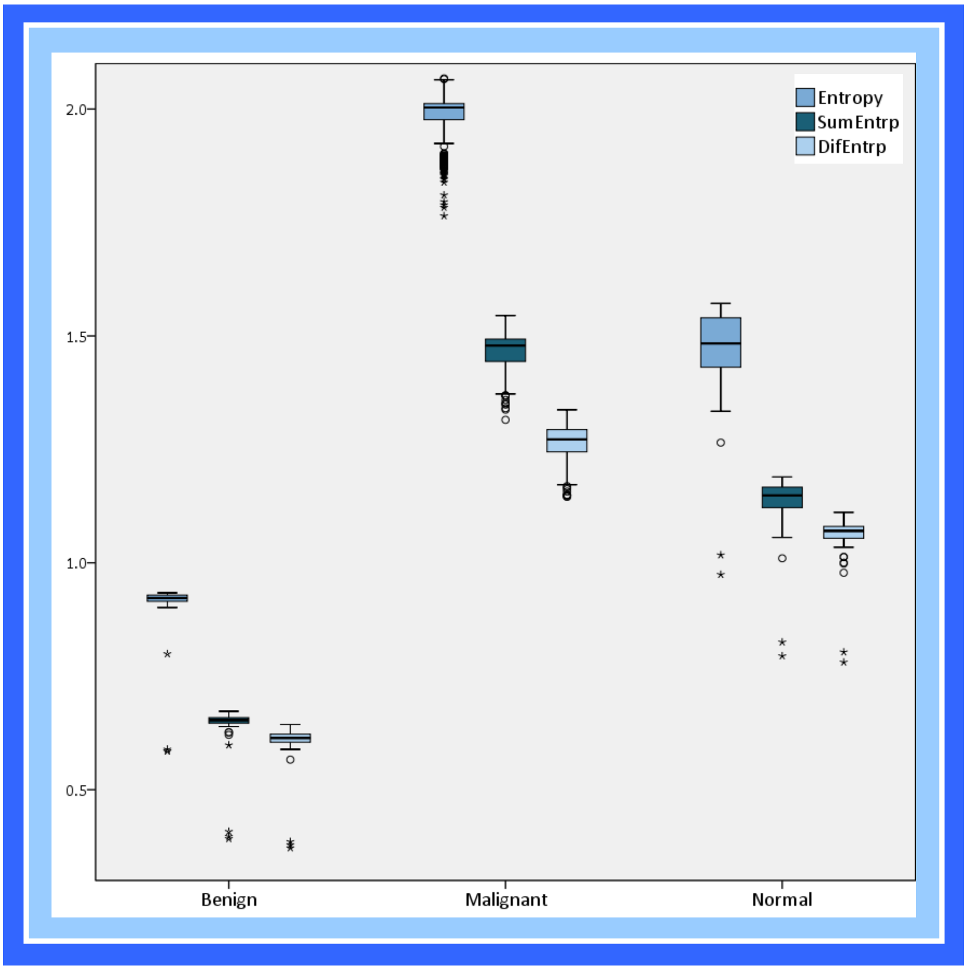


Figure 4.34- Box-whisker plot demonstrating entropy features as measured in normal, benign and malignant tissue

4.3.6 Classifications of Histological Subtype and Grade

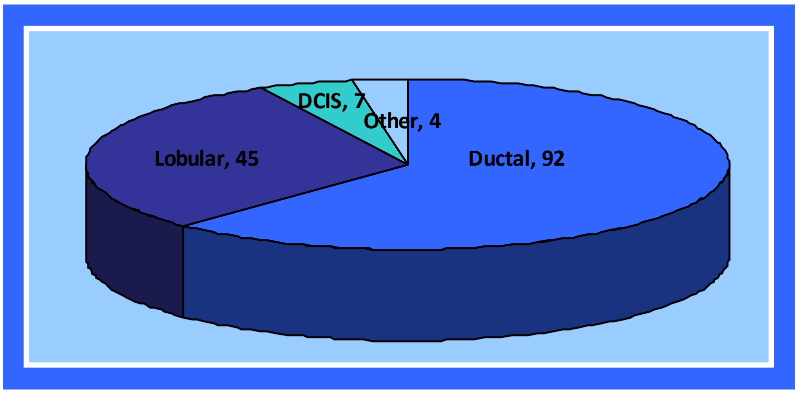


Figure 4.35- Split of lesion histopathology included in classification of histological subtype study

Histological subtype was extracted from the pathological report obtained from the pre-MRI biopsy. Histological subtyping information was available for all lesions and the data used is summarised in Figure 4.35.

Classification of the data was performed both pair-wise and also across the entire dataset. The results of the classification when using all COM features and only entropy features are presented in Table 4.20. Classification accuracy was good for the whole dataset with an overall accuracy of 74.7% and ROC area of 0.816.

			All COM features		*All Entropy features*	
			Classification Accuracy	*ROC*	*Classification Accuracy*	*ROC*
Ductal	vs.	Lobular	71.4 %	0.745	64.7 %	0.632
Ductal	vs.	DCIS	96.8 %	0.781	96.8 %	0.433
Lobular	vs.	DCIS	95.1 %	0.772	93.2 %	0.616
All Histological Subtypes			74.7 %	0.816	64.3 %	0.637

Table 4.20- Summary of histological classification accuracies and ROC areas for pair-wise and global classification using k-NN (k=3) and 10-fold cross validation

Data is displayed graphically in Figure 4.36 for the entropy features which demonstrate the relative spread of the lobular entropy measurements compared with either the ductal cancers or DCIS.

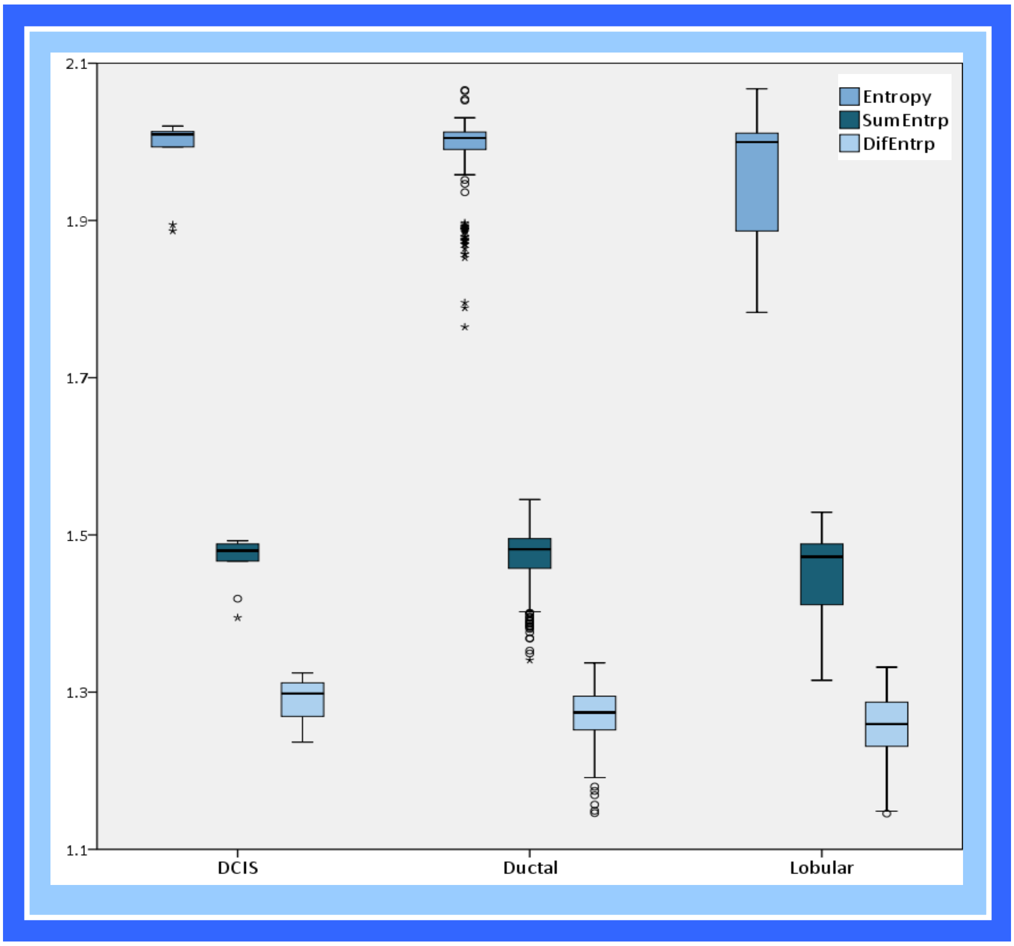


Figure 4.36- Summary of entropy-based features for histological subtypes of breast cancer

Out of the 220 COM features calculated, there were found to be 114 significantly different ($p<0.05$; Kruskal Wallis), and the summary of contributions from each individual feature category is presented in Figure 4.37.

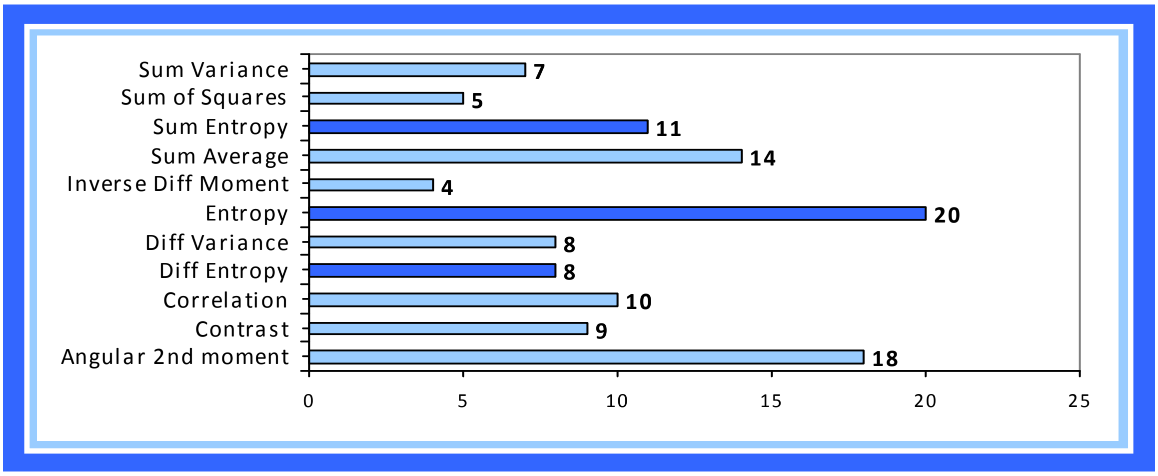


Figure 4.37- Number of significantly different features in each COM group between histological subtypes (p<0.05; Kruskal Wallis). Entropy features are highlighted in dark blue. There is a maximum of 20 features possible for each texture parameter.

Average values for each individual COM feature were also considered, and statistical analysis performed between groups using a Kruskal-Wallis test to identify features significantly different between the histological subtypes. Results are presented in Table 4.21.

	Ductal	*Lobular*	*DCIS*	*Kruskal-Wallis*
Angular 2nd Moment	0.0120±0.002 (0.010±0.01)	0.013±0.002 (0.010±0.01)	0.012±0.002 (0.0110±0.01)	p=0.001
Contrast	185±20 (184±100)	187±20 (187±100)	205±10 (206±50)	p=0.005
Correlation	0.080±0.06 (0.080±0.3)	0.0800±0.07 (0.070±0.4)	0.030±0.04 (0.210±0.1)	p=0.021
Difference Entropy	1.27±0.04 (1.27±0.2)	1.26±0.04 (1.26±0.2)	1.29±0.03 (1.30±0.09)	p=0.001
Difference Variance	66.0±8 (65.5±50)	64±8 (64±50)	73.0±7 (74.8±20)	p=0.013
Entropy	1.99±0.05 (2.00±0.3)	1.96±0.07 (2.00±0.3)	1.99±0.05 (2.01±0.1)	p=0.002
Inverse Diff Moment	0.100±0.01 (0.090±0.08)	0.0990±0.02 (0.09±0.1)	0.0880±0.007 (0.0890±0.02)	p=0.124
Sum Average	66.0±1 (65.7±5)	66±1 (66.1±5)	65.0±1 (64.7±3)	p<0.001
Sum Entropy	1.47±0.04 (1.48±0.20)	1.45±0.05 (1.47±0.2)	1.47±0.03 (1.48±0.1)	p=0.001
Sum of Squares	100±8 (102±40)	100±6 (101±40)	106±5 (106±10)	p=0.070
Sum Variance	218±20 (219±100)	216±20 (215±100)	220±10 (224±10)	p=0.340

Table 4.21- Summary of average ± standard deviation (median ± range) of each COM feature for each histological subtype and the results of Kruskal Wallis statistical testing, with significant differences highlighted in grey (p<0.05; Kruskal Wallis). All mean/ median values are quoted to 3 significant figures (3 sigfigs) and standard deviation/ ranges to 1 sigfig.

	Grade 1	*Grade 2*	*Grade 3*
Ductal	1	30	61
Lobular	-	35	10

Table 4.22- Summary of histological grade for ductal and lobular cancers

From these it is clear that there are significant differences in most features between each histological subtype, with only inverse difference moment, sum of squares and sum variance exhibiting no significant differences.

As there were only 7 DCIS cases, these were not considered for assessing histological grade. The histological grade of the ductal and lobular cancers is summarised in Table 4.22. Grade 1 ductal cancers were not considered further since only one lesion was identified - most likely due to the minimum size criteria. A summary of classification results for ductal and lobular histological grade is provided in Table 4.23. Classification accuracy was excellent using all features, with high values of ROC obtained when classifying between grades two and three. For the lobular cancers, entropy features alone also resulted in excellent classification of the cancers, suggesting that these features may dominate in the textural description of lobular cancer grade.

		All COM features		*All Entropy features*	
		Classification Accuracy	*ROC*	*Classification Accuracy*	*ROC*
Ductal	Gd2 vs. Gd3	71.8 %	0.811	60.4 %	0.631
Lobular	Gd2 vs. Gd3	82.6 %	0.886	78.8 %	0.738

Table 4.23- Summary of classification results for histological grade of ductal and lobular cancers

4.3.7 Classifications of Immunohistochemical Subtype

The receptor status of the 148 lesions was identified using the final pathological biopsy report. HER2 status was unavailable for one lesion therefore to distinguish between the HER2 and triple negative breast cancers (TNBC) there were only 147 lesions considered. The split in each category is as shown in Table 4.24.

	Lesion Population		***Classification Results***			
			All COM features		*All Entropy features*	
	Positive	*Negative*	*Classification Accuracy*	*ROC*	*Classification Accuracy*	*ROC*
ER Status	105	43	73.4 %	0.776	68.8 %	0.643
HER2 Status	50	97	68.3 %	0.733	62.4 %	0.607
TN Status	26	121	81.1 %	0.801	79.0 %	0.618

Table 4.24- Number of lesions with each immunohistochemical subtype and classification results

The TNBC are well differentiated from the rest of the cancers with a high classification accuracy of over 80% and excellent ROC area of 0.801. This suggests underlying grey-level patterns that can be well differentiated between the two types of cancers. The ER+/- cancers also have good classification accuracy with reasonable ROC areas. HER2+/- cancers seem to be less well distinguished using texture analysis, with overall lower accuracies but maintaining good ROC areas.

The number of significantly different features was calculated for each receptor status for each immunohistochemical subtype (i.e. ER+/-; HER2+/-; TN/Other) using a Mann Whitney U test. There were 24 out of the total 220 COM features significantly different for both triple negative and HER2 categories. For the ER subtype, there were 71 significantly different features between ER+ and ER- cancers. The breakdown of these features into COM categories is shown in Figure 4.38.

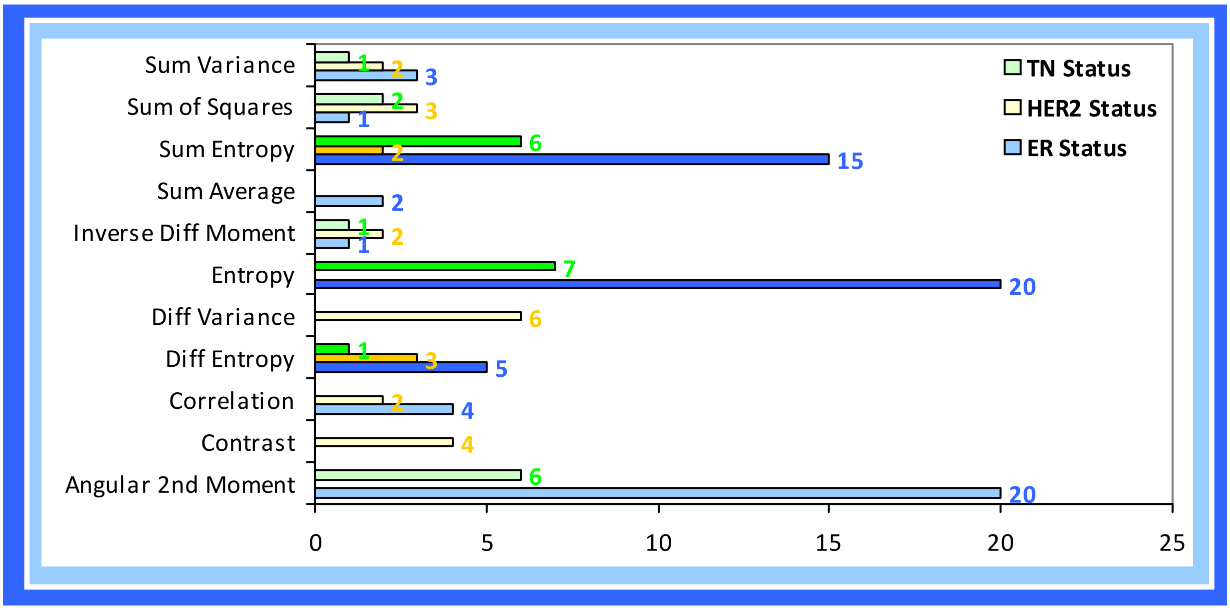


Figure 4.38- Number of significantly different features between receptor statuses from each COM category, for each immunohistochemical status category ($p<0.05$; Mann Whitney U). Entropy features are highlighted in darker shades. There is a maximum of 20 features possible for each texture parameter.

When statistical comparisons were carried out using a Mann-Whitney U test on the absolute values for each COM category:

- ER subtypes demonstrated significant differences in entropy ($p<0.001$), sum entropy ($p<0.001$) and sum variance ($p=0.011$) features.
- HER2 subtypes demonstrated no significant differences between any of the COM feature absolute values.
- TN demonstrated significantly different values from other cancers in the angular second moment ($p=0.005$), entropy ($p=0.003$), sum entropy ($p=0.007$) and sum variance ($p=0.018$) features.

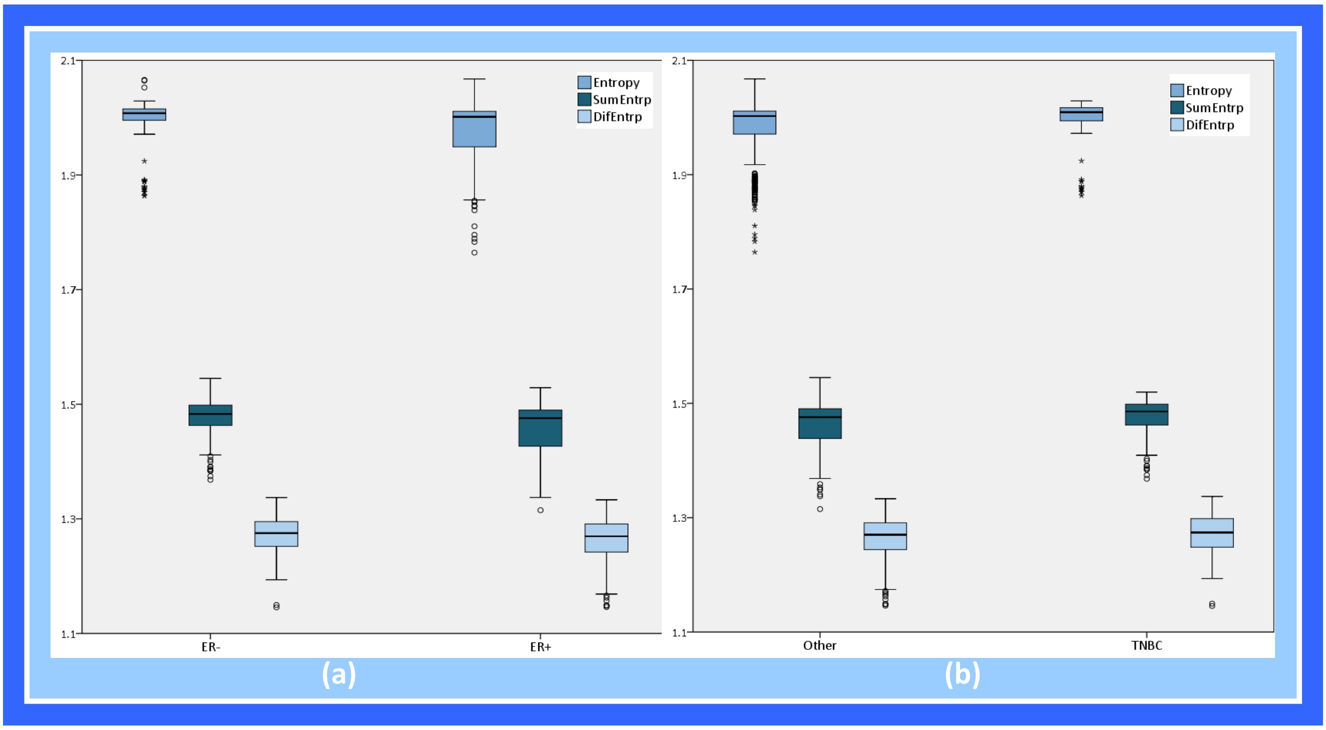


Figure 4.39- Entropy box-plots for both ER (a) and TN (b) immunohistochemical subtypes of breast cancer

4.3.8 Classifications by Molecular Subtype

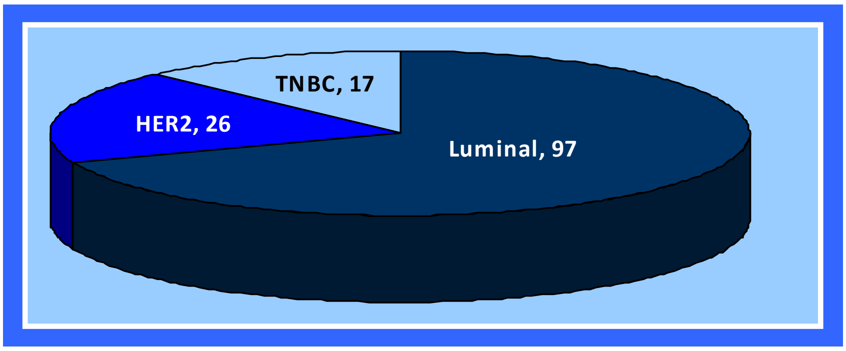


Figure 4.40- Breakdown of lesions into molecular subtypes

Only invasive cancers were considered for molecular subtype categorisation. As HER2 status was unavailable for one lesion and there were 7 cases of DCIS, only 140 cases were included in this study as shown in Figure 4.40.

Classification was performed between the molecular subtypes which resulted in a classification accuracy of 74% and ROC of 0.780 (Table 4.25). Classification using only entropy-based features also resulted in a reasonable classification accuracy of 65%, although ROC areas were lower at 0.624.

	All COM features		*All Entropy features*	
	Classification Accuracy	*ROC*	*Classification Accuracy*	*ROC*
Luminal vs. HER2 vs. TNBC	73.0 %	0.783	65.2 %	0.625

Table 4.25- Classification results for molecular subtypes.

Results for this are presented graphically in Figure 4.42. From this plot, the slight increased value in entropy features for both the triple negative cancers and HER2 cancers can be seen relative to all other luminal cancers. The spread of entropy features is greater for the luminal cancers compared with either the HER2 or TN-type cancers.

Statistical evaluation of the raw texture feature values was also carried out and the number of significantly different features for each COM category is presented below in Figure 4.41. Again, it is clear from this plot that entropy, sum entropy and angular second moment features appear to be most significant at characterising differences in texture between these molecular sub-groups.

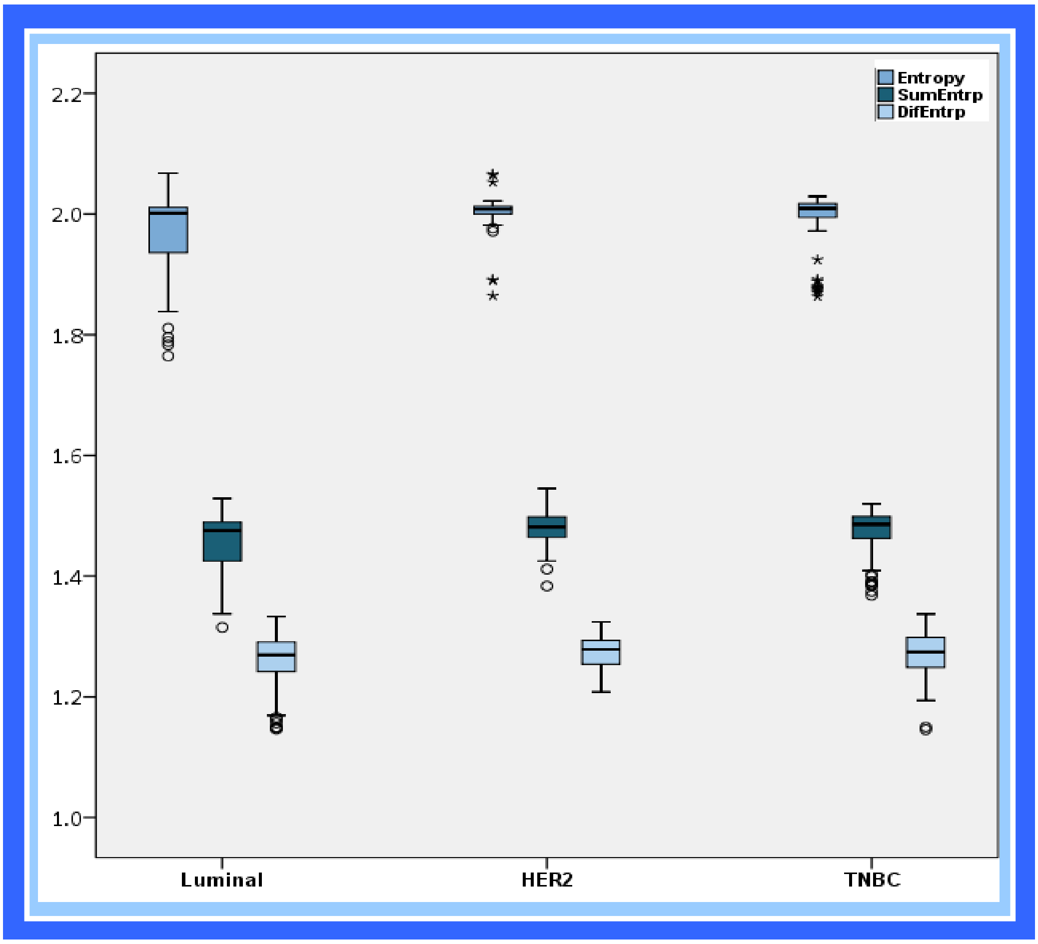


Figure 4.42- Graphical box-whisker plot showing entropy distributions for each molecular subtype of breast cancer.

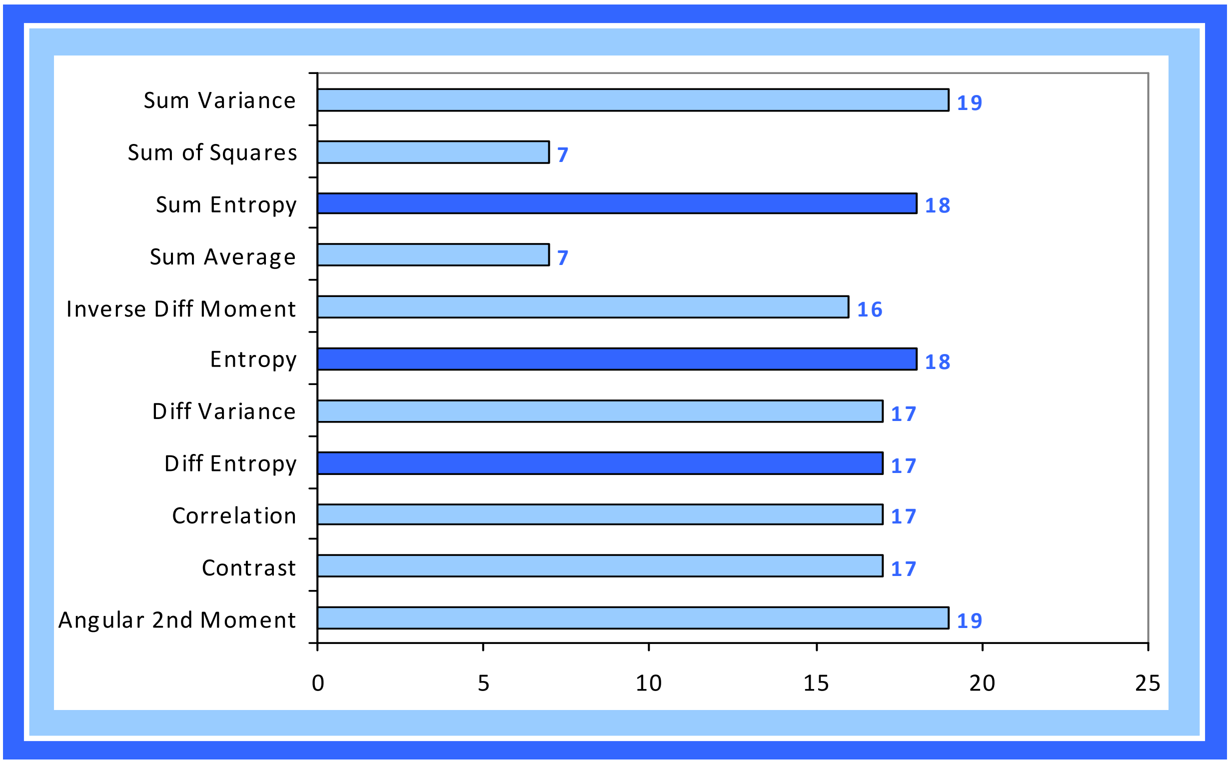


Figure 4.41- Number of significantly different features in each COM category when all molecular subtypes of breast cancer are considered ($p<0.05$; Kruskal Wallis test). Entropy features are highlighted in darker shades. There is a maximum of 20 features possible for each texture parameter.

Average feature values for each of the categories, and the findings from the Kruskal Wallis test between each molecular subtypes are presented in Table 4.26.

	Luminal	*HER2*	*TNBC*	*Kruskal-Wallis*
Angular 2nd Moment	0.0121±0.002 (0.0110±0.01)	0.0111±0.001 (0.0109±0.01)	0.0115±0.002 (0.0108±0.01)	p<0.001
Contrast	186±20 (187±100)	182±20 (182±80)	187±20 (186±100)	p=0.265
Correlation	0.0771±0.06 (0.0725±0.3)	0.0968±0.06 (0.0863±0.3)	0.0877±0.07 (0.0771±0.3)	p=0.197
Difference Entropy	1.26±0.04 (1.27±0.2)	1.27±0.03 (1.28±0.1)	1.27±0.04 (1.274±0.2)	p<0.001
Difference Variance	66.0±8 (65.6±49)	65.3±8 (65.7±35)	64.9±9 (64.9±40)	p=0.358
Entropy	1.97±0.06 (2.00±0.30)	2.00±0.03 (2.01±0.2)	1.99±0.05 (2.01±0.2)	p<0.001
Inverse Diff Moment	0.0975±0.02 (0.0950±0.1)	0.0954±0.01 (0.0935±0.04)	0.0945±0.02 (0.0914±0.1)	p=0.013
Sum Average	65.8±1 (65.9±6)	65.8±1 (65.7±5)	65.8±1 (65.8±4)	p=0.260
Sum Entropy	1.46±0.04 (1.48±0.2)	1.48±0.03 (1.48±0.2)	1.47±0.04 (1.49±0.2)	p<0.001
Sum of Squares	101±8 (102±40)	100±9 (102±40)	102±9 (103±40)	p<0.001
Sum Variance	216±19 (215±100)	220±20 (224±100)	221±20 (2223±100)	p=0.055

Table 4.26- Summary of average ± standard deviation (median ± range) of each COM feature for luminal, HER2 and TNBC molecular subtypes and the results of Kruskal Wallis statistical testing, with significant differences highlighted in grey ($p<0.05$; Kruskal Wallis). All mean/ median values are quoted to 3 significant figures (3 sigfigs) and standard deviation/ ranges to 1 sigfig.

These results demonstrate the importance of the entropy, sum entropy and angular second moment features in discrimination between the molecular subtypes of breast cancer.

Pair-wise classification of the molecular subtypes of cancer was then performed with results presented in Table 4.27. Statistical evaluation of the entropy features was also performed for the pair-wise classification with significantly different entropy features highlighted in the table (p<0.05; Mann Whitney U).

			All COM features Cross Validation		*Entropy Features Cross Validation*		*Statistical comparison of entropy features*
			Classification Accuracy	ROC Area	Classification Accuracy	ROC Area	
Luminal	vs.	HER2	84.6 %	0.717	84.0 %	0.663	p=0.003
Luminal	vs.	TNBC	78.3 %	0.797	74.9 %	0.623	p=0.001
HER2	vs.	TNBC	68.8 %	0.750	62.5 %	0.625	p=0.822

Table 4.27- Pair-wise classification of molecular subtypes for all COM features and entropy features only. Statistical evaluation of entropy features is also presented, with p<0.05 considered significant (Mann Whitney U)

The results presented highlight that significantly different entropy features are demonstrated for luminal cancers compared with the higher grade, more aggressive HER2 and TN cancers. These trends were generally mirrored by the results from the classification with those comparisons resulting in good classification accuracies and ROC areas.

There were, however, no significant differences between the TNBC and HER2 entropy features.

4.3.9 Signal to Noise

Signal to noise measurements were calculated for all lesions included in this study and compared between the histological and molecular subtypes of cancers as presented below in Figure 4.43. Error bars depict the standard deviation of the SNR measures in each subtype category. The large standard deviations are likely to be attributable to the distribution of scans performed on 1.5T and 3.0T resulting in contrast enhancement characteristics differing slightly between the different field strengths. While this should have minimal impact on texture analysis outcome (due to the same pixel size and image normalisation being performed), the absolute signal intensity values were used for calculation of SNR and therefore this is likely to increase the range of values encountered. Statistical pair-wise comparisons were performed between each histological and molecular subtype using a students t-test, and one way ANOVA across all subtypes, and results of statistical analysis are presented in Table 4.28. There were found to be no significant differences in measured SNR values in each pair-comparison, suggesting that differences in texture analysis were unlikely to be attributable to this particular factor.

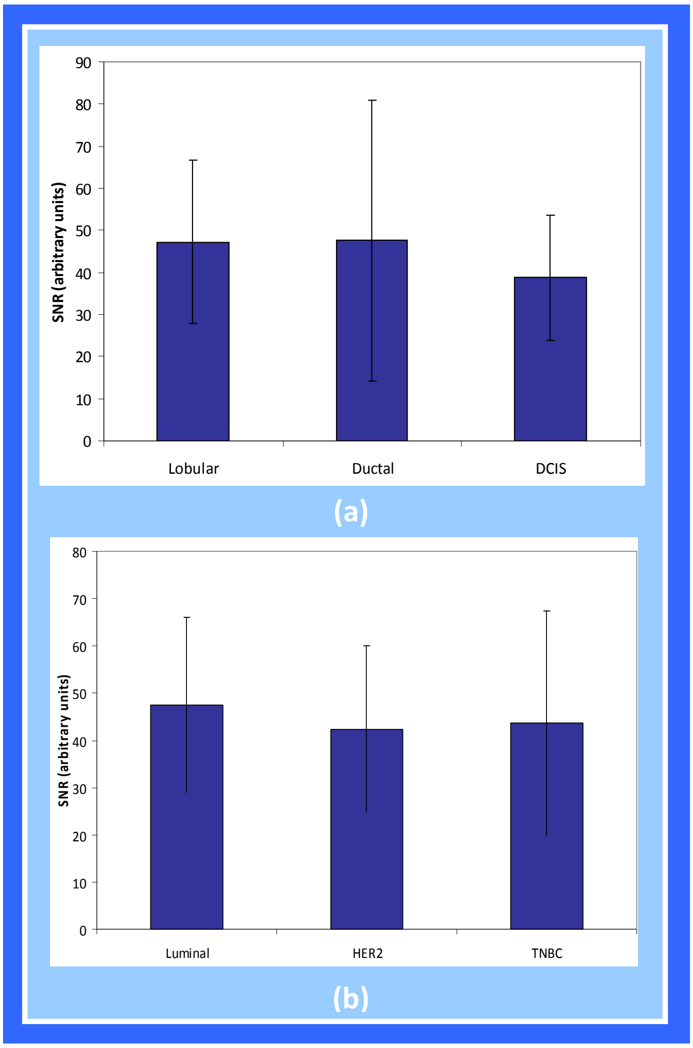


Figure 4.43- Summary of signal to noise measurements for (a) histological subtype groups and (b) molecular subtype groups

Histological Subtype Comparison	*SNR comparison p-value*	*Molecular Subtype Comparison*	*SNR comparison p-value*
Ductal vs. Lobular	p=0.951	Luminal vs. HER2	p=0.443
Ductal vs. DCIS	p=0.601	Luminal vs. TNBC	p=548
Lobular vs. DCIS	p=0.396	HER2 vs. TNBC	p=0.856
Histological subtype	p=0.736*	Molecular subtype	p=0.163*

Table 4.28- Summary of statistics performed on SNR differences between different histological and molecular subtypes of cancer. There were no significant differences ($p<0.05$; students t-test; *$p<0.05$; ANOVA)

4.4 DISCUSSION

This work has demonstrated the potential clinical utility of texture analysis and outlined the uses of the techniques in various applications, including identification of malignancy, differentiation from benign disease and classification of breast cancer subtypes.

All the results have been reported in terms of the percentage of correctly classified data-points as well as area under the ROC curve. This provides an assessment of how well separated the classes are and is not dependent on any decision threshold that may be selected. As well as classifying the data in feature space, statistical assessment of the raw feature values has been performed in order to quantify whether the values themselves can be used in differentiating between breast cancer subtypes.

Reproducibility assessment was performed and demonstrated to be quite variable across all the COM parameters; however, the majority of features demonstrated a CoV of less than 25% across both intra- and inter-reproducibility measurements. The entropy based features were also demonstrated to be consistent across both readers.

Classification of malignant from normal tissue resulted in near perfect classification with an accuracy of 97.9% (ROC=0.998), and this classification result was slightly improved when only entropy features were used (99.8%; ROC=0.998), suggesting that these features differ significantly between normal and malignant tissue.

With the addition of benign lesions, the classification was found to be perfect, with a classification accuracy of 100% (ROC=1.000), and significant differences between the three tissue types when considering only the entropy features ($p<0.001$; Kruskal Wallis). The entropy features were found to be lowest for benign tissue and highest for malignant tissue. This would indicate a more heterogeneous nature of the malignant tissue compared with either of the other two tissue types, and suggests that benign lesions may actually demonstrate a relatively homogeneous pixel intensity profile, which corresponds with the BIRADS-2 MRI descriptor that these types of lesions generally have a particularly homogeneous appearance on MRI [53].

The cancers were then split into various subtypes and histological grading where TA was found to perform well. While accuracies over 90% were achieved for the DCIS comparisons with both ductal and lobular cancers, this could be attributable to the relatively low numbers of these cases compared with the number of ductal and lobular cancers included. However, ductal and lobular classification was good with accuracy of over 70% and corresponding ROC areas of 0.745. When only entropy features were used, these were found to provide a lower accuracy at 65% (ROC=0.632). Significant differences were found between most of the COM features between these three cancer types ($p<0.05$; Kruskal Wallis). Entropy features demonstrated higher values for the ductal cancers (both in situ and invasive) compared with the lobular lesions.

When the histological grading of both lobular and ductal cancers was considered, the classification accuracy was found to be good, with accuracies of over 70% and ROC areas greater than 0.8. When only entropy features were considered, the lobular grade 2 and 3 cancers were still well classified (78%; 0.738), while the accuracy for ductal was lower (60%; 0.631). This would tend to suggest that the histological grade of cancers is reasonably well represented by the entropy textural features, particularly for the lobular cancers.

Classification of the lesions according to the immunohistochemical receptor status was then performed to identify if differences in whether these receptors genes were present had an influence on appearance in MR images.

Classification was found to be good for both ER status (73.4%; 0.776) and for triple negative cancers (81.1%; 0.807), however was found to be significantly poorer for classification between HER2 groups (68.3%; 0.733).

Both ER and TN groups demonstrated significant differences in the raw feature values between positive and negative state in entropy features, although the classification accuracies were lower when only entropy features were considered (ER: 68.8%, 0.643; TN: 79.0%, 0.618).

The triple negative and ER- cancers appeared to demonstrate overall higher values of entropy, difference entropy and sum entropy compared with the other and ER+ cancers, suggesting a more heterogeneous pixel profile.

Full molecular descriptions of the cancers were then considered, by categorising into Luminal, HER2 and TNBC subtypes. Classifications remained good with accuracies of 73.0 % and ROC areas of 0.783. Results for the entropy features resulted in disappointing results (65.2%; 0.625).

Statistical comparison of the categories demonstrated significant differences in the angular second moment, entropy and sum entropy parameters, and entropy was found to be higher in both the HER2 and TNBC lesions compared with the luminal categories ($p<0.05$; Kruskal Wallis).

Pair-wise classification and statistical evaluation of all molecular subtypes demonstrates that the luminal categories appear to be most distinct from HER2 and TNBC subtypes, in keeping with their different biological behavious (such as the way they are treated and time-frames to recurrence etc.)

Signal to noise parameters were also assessed to ensure that classification accuracies and statistical evaluations were not influenced by differences in SNR measures between each group. No significant differences were found between any of the histological or molecular subtypes ($p>0.05$; ANOVA).

We have demonstrated the entropy features appear to consistently be the most discriminating between different regimens of classifying breast cancer. This is in agreement with findings from other groups that entropy can be used in the differentiation between malignant and benign lesions [103, 104] and different histological subtypes of cancers [110]. The entropy features reflect internal pixel distribution patterns, which could potentially link directly or indirectly with underlying growth patterns. Therefore TA could provide a method of mapping intra-tumoural heterogeneity and provide a useful means of classifying between different subtypes of cancers.

Most previous studies in which TA has been applied in the field of breast MRI have been in the identification of malignancy within healthy breast tissue and in the discrimination between malignant and benign lesions [103-105, 109]. Further characterisation of such lesions is required for clinically useful treatment planning, monitoring and distinguishing therapeutic effects [21, 150]. Whilst identification of a lesion using TA methods requires differences to be detected between suspicious and surrounding areas of tissue, characterisation is a more complex and challenging area as the difference in texture between pathologies needs to be clarified [109]. The versatility and applicability of TA in the diagnosis, classification and grading of cancers in a clinical setting is an emerging field and recent studies have used TA in differentiation between invasive cancer and DCIS on mammographic images [109], classification of different types of glioblastoma [150] and distinguishing between gliomas and metastases [95] on MR images. We have taken this one step further by aiming to classify cancer types based on underlying molecular subtype descriptions to identify any differences in growth patterns that may not be visually apparent.

It should be noted that there appear to be no other imaging modalities capable of discriminating between subtypes of cancer reliably. In a study carried out by Berg et al., it was found that mammography plus ultrasound combined resulted in an ROC area of 0.9 when predicting the probability of malignancy, compared with mammography alone where the ROC area was significantly lower at 0.74-0.78 [151, 152]. We have demonstrated that texture analysis can differentiate between malignancy, normal and benign tissue with an ROC area of 1.000 (perfect classification) and therefore this technique appears to outperform traditional breast imaging assessment.

Our results demonstrate that luminal cancers appear to have lower entropy feature values compared with the triple negative and HER2 cancer groups, and that entropy features are the most significant features in differentiating between these categories. The growth patterns of luminal cells generally exhibit tight cell-cell junctions with epithelial-differentiated monoloayers [23]. Triple negative cancers, however, demonstrate a more scattered morphology compared with the luminal cancers and exhibit a solid architecture and necrosis which is morphological distinct compared with the luminal tumours [22]. There have also been reports in the literature that there is an overlap in the histological growth patterns of triple-negative and HER2 over expressing cancer subtypes [22], which appear to correlate with our findings. It would be worthwhile to further investigate the correlation of texture features as calculated using MR images with the indicators of systemic spread, such as nodal status and vascular invasion, although this study is outwith the scope of this work.

While it is recognised that texture analysis does not represent the underlying growth pattern directly, due to the different orders of magnitude in resolution, it is possible that there is an indirect link between the growth patterns and resulting pixel distributions on MR imaging, which may or may not be due to patterns of contrast uptake, or relative relaxation time values. Such characterisation of high and low grade gliomas based on entropy features has been reported for computed tomography (CT) imaging [153] and imaging the heterogeneity of lesions is becoming more topical [85, 106-108] in the drive to fully characterise entire lesions rather than small sub-samples as obtained using biopsy.

The main limitation in this study is in the subtyping of breast cancers with such a small sample size. While this population represents a group larger than similar studies reported in the literature [103, 104, 110], the number of subgroups that data was categorised into requires a large number of data samples in order to increase classification accuracies and statistical certainty. We have demonstrated promising initial findings, however in order to fully characterise and classify lesions based only on the entropy features, which we believe should be possible, it is likely that a large-scale cohort of patients would be required- most probably associated with a multi-centre approach. It should be noted, however, that the cohort of patients considered in this study is not representative of breast cancer patients as a whole, but rather the typical subgroup that are routinely referred for MRI examination and therefore this does not necessarily translate as a technique that can be widely applied to all breast cancer patients as a whole. Such a study would require considerable escalation of patient numbers and referral criteria.

In the histological classification of data few DCIS lesions were available for comparison with invasive cancers, reflecting the reality of utilisation of MRI in clinical practice for DCIS and invasive breast cancer. It is recognised that invasive cancer may coexist with DCIS [109] and therefore comparisons may be complex. However in each case the main tumour body was used for placement of the ROI in accordance with the radiologist report and due consideration was given to final pathology.

Data was utilised in this study that was acquired under two different imaging conditions, which may attribute another variable to the study. The rationale for this was to provide adequate patient numbers for the research, however this does also potentially represent the clinical reality where more than one MRI system may be used for a clinical service. This also increases weight to the findings from our study, as we have demonstrated that good results can still be obtained despite using different equipment across the entire patient cohort. There were no significant differences in the measured SNR associated with the lesions included in the study and therefore there were no attributable effects from this particular parameter. We also demonstrated in Chapter 3 that robust texture analysis can be performed across different magnets, and clinical protocols, providing the spatial resolution remains constant, as it does in this study.

We also utilised a highly specific analysis technique with only 3 slices included in the texture analysis and only lesions greater than a minimum size of 8mm. Texture classification results are known to depend on the number of data-points included in the analysis and therefore to include the entire number of imaging slices for every lesion included in the study could have introduced an artificial dependence on the size of the lesion, with larger lesions contributing more data-points than smaller lesions. Also, luminal cancers are generally smaller than TNBC and therefore direct comparisons between the two groups would not have been possible without some correction. We therefore chose a method that would provide the same number of datapoints from every lesion. Only lesions greater than 8mm were included in the analysis. This was again to ensure consistency of analysis between lesions- a fixed size of the ROI was used of 10×10, which corresponded with the reported lesion size. This provides a major limitation to our study as we are unable to assess as to whether a minimum lesion size limit may exist for texture analysis to be useful in the clinic: rather, we hypothesise that this is most likely to be pixel-size limited.

The data presented provides a quantitative method for assessing breast lesions using MR imaging. Such non-invasive methods have potential in patients where sampling (multiple) breast lesions is anatomically difficult or could lead to histological underestimation of the disease present [95, 109]. Therefore it is concluded that TA could offer a real option for lesion characterisation in routine clinical practice.

CHAPTER SUMMARY: This chapter has demonstrated that texture analysis can be used with excellent classification between normal, benign and malignant tissue. Classification accuracies are good between each method of subtyping the cancers, with the exception of HER2 status. Entropy features appear to be significantly different between these categories, however do not result in high classification accuracies when used alone for subtype classification. The entropy features, which represent heterogeneity of pixel distributions within a region of interest, appear to follow the expected trends associated with underlying growth patterns (e.g. more entropy in high grade triple negative cancers), so there may potentially be an indirect link between the pixel distributions and/or contrast uptake on MR images and the underlying histological growth pattern, although the mechanism for this is unclear.

CHAPTER FIVE: COMPARISON OF TEXTURE ANALYSIS TECHNIQUES

OUTLINE OF CHAPTER: The previous chapter outlined the use of texture analysis in a clinical cohort of patients, and the usefulness in identification and classification of breast cancers. This chapter considers a different method of texture analysis, known as the max-min technique, to reanalyse the work performed in Chapter 4. The aim of this chapter is to carry out a technical comparison of different TA techniques which will either reinforce the method or identify if the outcome is dependent on the technique used.

5.1 INTRODUCTION

The co-occurrence matrix has been widely reported as a useful model in texture analysis for differentiating between tissue types in MR images. Whilst this technique utilises second order statistics [87, 91, 93, 95, 96, 103-105, 109, 110], first order statistics have also been demonstrated to provide a level of discrimination between different textures in medical imaging [154]. First order statistics, such as local extremes of signal intensity, are believed to most closely match what the eye perceives as 'texture' [86], and therefore the 'max-min' method for texture analysis was described in 1977 by Mitchell et al. as a method of texture analysis [122].

By using different degrees of smoothing, and calculating the number of extremes at each smoothing level, the ratios of these values can be used to provide a measure of image texture. This technique has been readily applied successfully in ultrasound, particularly for differentiation between various liver diseases [155, 156] and has the advantage of being computationally simple.

This study aimed to identify whether this simple first-order technique could be used in the differentiation of malignant and normal tissue, and in the classification of different cancer subtypes. This was performed on the same patient cohort as described in the previous chapter to identify if results were consistent between the two methods or if they provided conflicting results.

5.2 MATERIALS AND METHODS

5.2.1 Patient Population

The patient population that this study was carried out on is the same as described in Chapter 4.2.1 and 4.2.2.

5.2.2 Texture Analysis

All regions of interest (ROI) were generated using MATLAB (v. 2012a, The Mathworks Inc.; Natick, MA) using the computer coding presented in Appendix B. The 10×10 square ROIs were placed in the same region and same slices as those described in the previous chapter (see Figure 5.44).

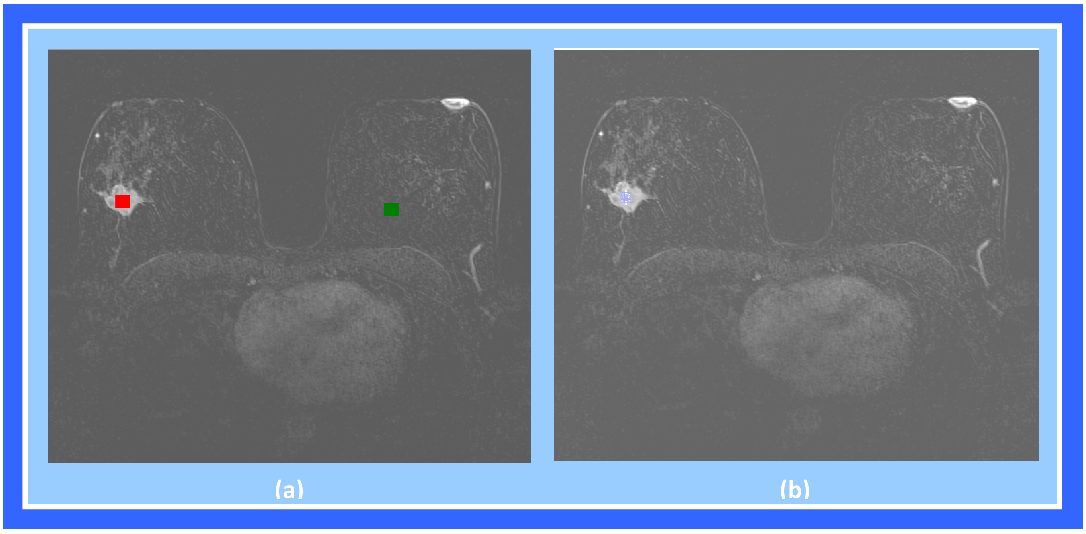


Figure 5.44- Diagram showing placement of ROI using MaZda (a) and a screenshot using the max-min method via MATLAB (b). Figure (a) also shows the contralateral ROI in normal tissue, this had to be placed separately using the max-min method.

Regions of interest of the same size were also placed in regions of non-enhancing contralateral tissue. Grey-level pixel values were stored in a matrix and exported as a text file. FORTRAN (Intel Corp.; Santa Clara, CA) code, developed in house using Xcode (Apple Inc.; Cupertino, CA) and run on the Mac OSX operating system, was then used for calculation of Max-Min texture parameters.

The grey-levels were represented by 9 bits/pixel (0-511) and the logarithm of grey level values was used to minimise the effect of brightness variations. Data was smoothed using the regime described by Mitchell [122] and in Chapter 2.2.2 (see Figure 2.7). A logarithmic threshold value of 0.1 was used and performed a total of five times, with an incremental value of 0.01. The number of extreme values detected for each smoothing process was expressed as a ratio to the first count of extreme values to ensure independence from the size of the ROI.

For each ROI, a total of four texture parameters were calculated, corresponding to the five smoothing values.

5.2.3 Statistical Analysis

The raw texture parameters calculated from the max-min algorithm were exported for statistical analysis using SPSS. A p-value less than 0.05 was considered to be significant. All comparisons as described in Chapter 4.2.5 were performed using this method- i.e. malignant vs. normal tissue and classifications of malignancy.

5.3 RESULTS

This study considered the same patient population as described in Chapter 4.3.1 and therefore 148 lesions were identified and included in this comparative study, with 147 lesions included in receptor and molecular subtype comparisons due to one HER2 status being unavailable.

5.3.1 Malignant vs. Normal

Statistical assessment of max-min texture parameters showed there was a significant difference between malignant and normal tissue ($p<0.001$, Mann Whitney U) and this is displayed graphically in Figure 5.45, with the malignant tissue demonstrating lower values of the measured tissue parameters compared with the normal tissue.

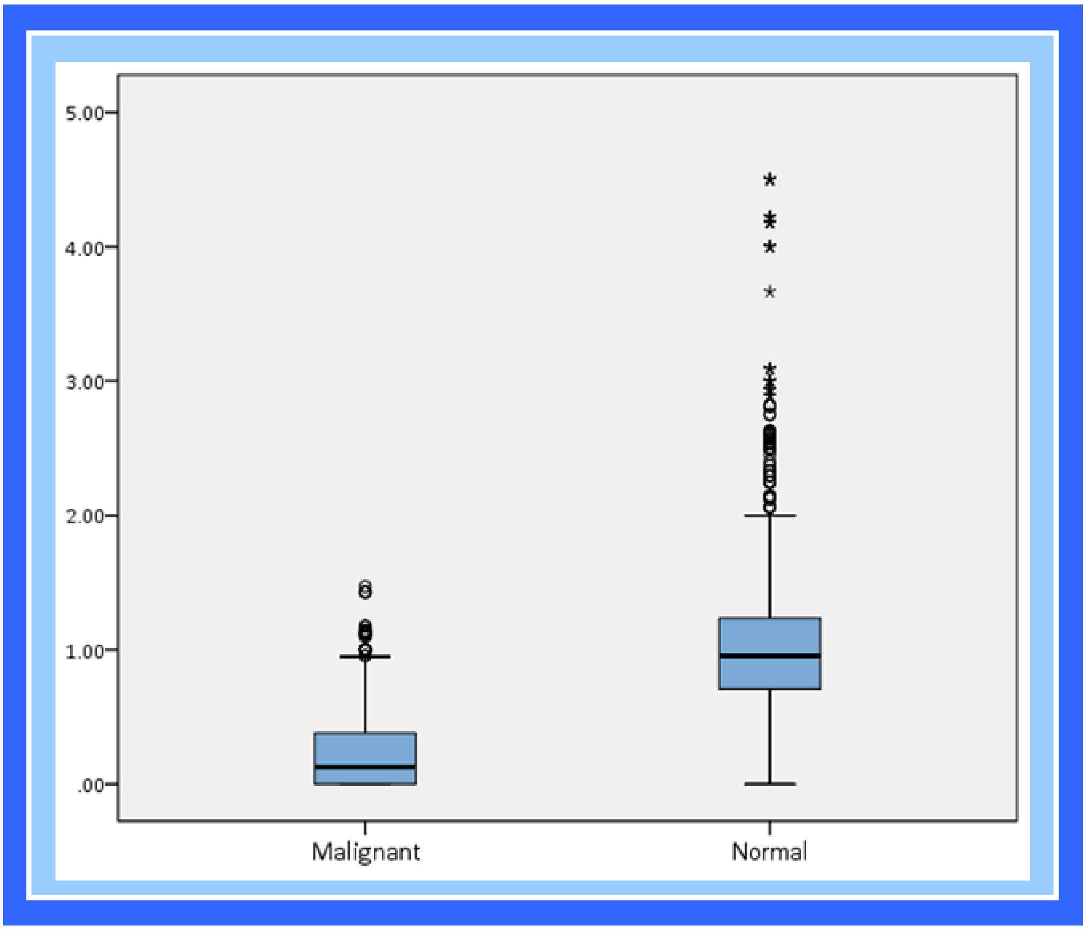


Figure 5.45- Box-Whisker plot showing distribution of average max-min texture parameters for malignant and normal tissue

5.3.2 Classification of Cancer Subtype

Data was then split according to the histological cancer subtype as described previously. Again, the difference between the calculated texture parameters was found to be significant ($p<0.001$; Kruskal Wallis) for comparisons of cancer subtype as demonstrated in Figure 5.46. Values of the max-min features are presented in Table 5.29. In pair-wise comparisons, statistically significant differences were found between DCIS and both lobular and ductal cancers ($p<0.05$; Mann-Whitney U), although no significant difference was demonstrated between ductal and lobular cancers ($p=0.533$).

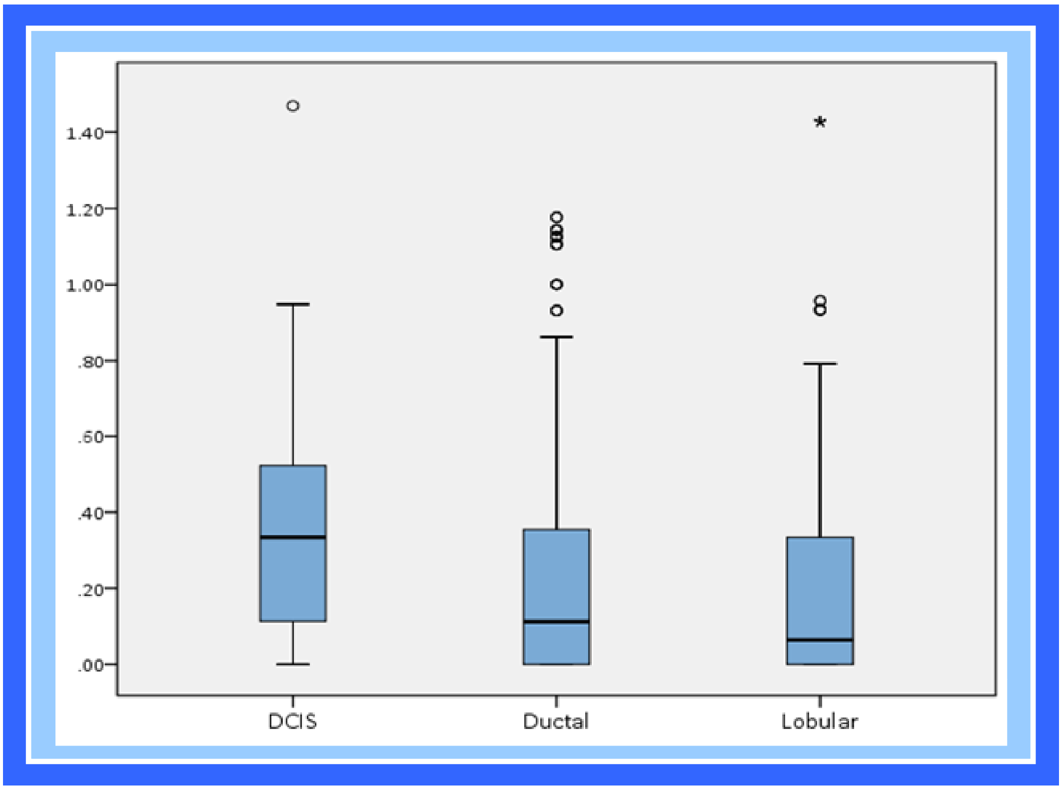


Figure 5.46- Distribution of average max-min texture parameters for each histological cancer

	Ductal	*Lobular*	*DCIS*
Mean ± std dev	0.22 ± 0.01	0.21 ± 0.01	0.39 ± 0.05
Median (min-max)	0.11 (0.00-0.18)	0.06 (0.00-1.43)	0.33 (0.00-1.47)

Table 5.29- Summary of mean, median and range of max-min values for each of the histological subtypes considered.

5.3.3 Classification of Receptor Status of Cancer

Comparisons of receptor status types demonstrated significant results for both the HER2 and triple negative types of breast cancer ($p<0.05$; Mann Whitney U). Results did not reach significance for comparison of ER+ and ER- types of breast cancer. Results of statistical comparisons are presented in Table 5.30.

	Mann Whitney U
ER Status	p=0.064
HER2 Status	p=0.041
TN Status	p<0.001

Table 5.30- Statistical results from comparisons of receptor statuses using max-min texture parameters

5.3.4 Classification of Molecular Subtypes

Average, median and range texture parameter values for each molecular subtype of cancer are presented in Table 5.31. Statistical comparison of texture features for molecular subtypes demonstrated significant differences ($p<0.001$, Kruskal Wallis) when all subtypes were considered together.

	Luminal	*HER2*	*TNBC*
Mean ± std dev	0.21 ± 0.01	0.24 ± 0.03	0.30 ± 0.02
Median (min-max)	0.12 (0.00-1.43)	0.04 (0.00-1.47)	0.25 (0.00-1.18)

Table 5.31- Summary of mean, median and range of max-min values for each of the molecular subtypes considered.

Pair-wise comparisons are presented in Table 5.32. Results are significant ($p<0.05$; Mann Whitney U) for comparisons of luminal and TNBC. Results are shown graphically in Figure 5.47

			Mann Whitney U
Luminal	*vs.*	HER2	p=0.203
Luminal	*vs.*	TNBC	p<0.001
HER2	*vs.*	TNBC	p=0.083

Table 5.32- Statistical findings from pairwise comparison of texture parameters for each molecular subtype

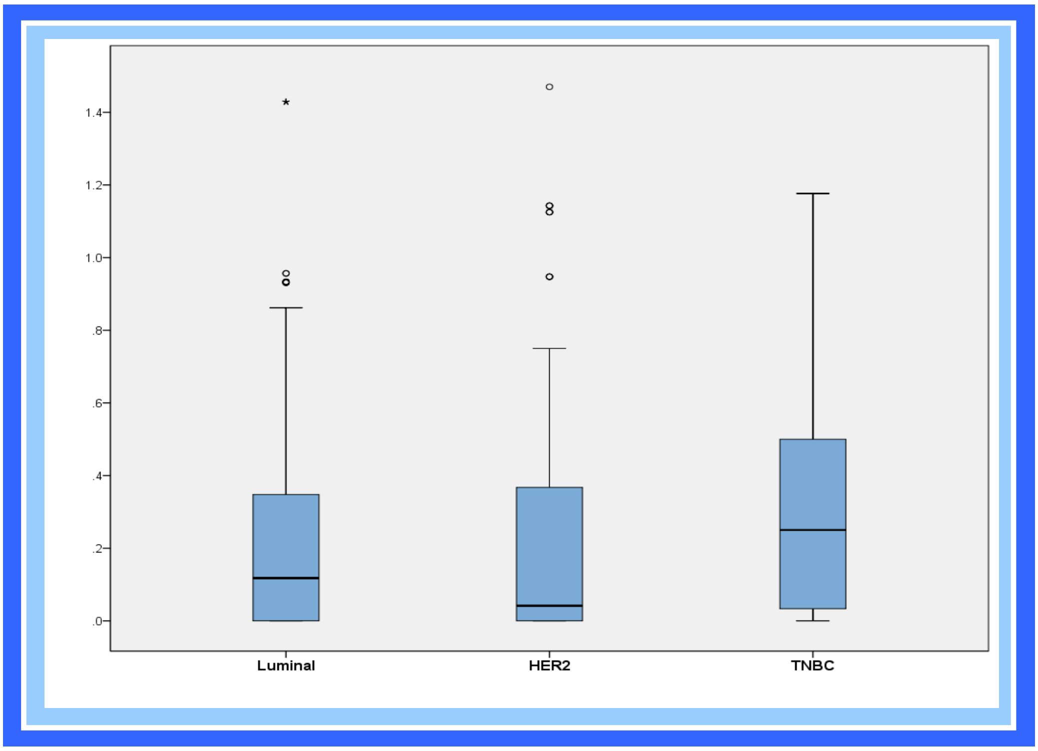


Figure 5.47- Distribution of average texture parameter calculated using the max-min technique for each molecular subtype of breast cancer

5.4 DISCUSSION

The max-min method of TA has demonstrated that it can provide good distinction between normal and malignant tissue and in the differentiation between cancer types.

In comparison between normal and malignant, there was a significant difference in the texture parameters calculated using the max-min method, with malignant tissue demonstrating lower parameter values compared with normal tissue. This would tend to suggest that malignant tissue has a greater number of extreme values that are not smoothed at higher threshold values and therefore resulted in overall lower values of the calculated texture parameters.

When considering histological subtypes of breast cancer, the calculated texture parameters were found to be significantly different ($p<0.05$; Kruskal Wallis), which is in agreement with the findings from the work presented in Chapter 4.

For molecular subtypes of cancer, significant differences were calculated between pair-wise comparisons of luminal and HER2 cancers with the average texture parameter largest for the TN cancers. This suggests a more heterogeneous texture within the lesion as characterised by a larger number of extreme values, resulting in lower calculated parameters using the max-min method. These results compare well with those demonstrated using the COM model and the entropy features, where both TNBC and HER2 were found to have the largest entropy values.

The results from this study confirm with those demonstrated in Chapter 4, suggesting that there are measurable texture differences on our MRI images of breast cancer that can be used to differentiate between normal and malignant tissue and also between different histological and molecular subtypes.

The texture parameters, as calculated using the max-min technique, appear to correlate most closely with the entropy features calculated using the COM model. While the entropy features provide a measure of disorder in the ROI, with larger values representing more heterogeneous lesions, the max-min technique is concerned with the number of minima and maxima detected in the ROI. As the texture parameters are calculated as a ratio after smoothing, those regions with more heterogeneous appearance (greater number of maximum and minimum extremes in grey level pixel value) will result in a lower value of the calculated texture parameter.

We therefore propose that the max-min technique also provides a measure of heterogeneity, with higher values of the calculated parameters representing more homogeneous lesions.

Whilst common first-order statistics such as mean, variance, skewness and kurtosis of pixel grey-level values are generally unable to differentiate between different textures in medical images, calculation of texture using the first-order statistical method implemented by the max-min algorithm has demonstrated an ability to differentiate between not only malignant and normal tissue but also cancer subtypes.

The advantage of the technique is that resulting features are essentially independent of the size of the ROI [157] and does not require the same computational power as more complicated methods. However, the method is only able to calculate one parameter, rather than the range provided by the COM model and therefore is unlikely to be able to model more complex image textures and is unlikely to be of use for prospective prediction of cancer subtype.

The main limitation of this study is the low number of DCIS cases included in the histological subtyping section, as in the previous chapter, however this was due to the selection criteria for patients referred for MRI examinations at our institution. We also did not consider the optimum threshold to use in differentiation between breast cancer subtypes but simply sought to replicate the work performed on the patient sub-group reported in Chapter 4.

The results using the max-min technique have therefore validated those that were obtained using the COM model, lending weight to the hypothesis that textural differences within the image that may not be visually apparent, contain the information within the pixel grey-level distribution to allow distinction between different subtypes of cancer. We hypothesise that the texture parameters as calculated using the max-min method provide a measure of the heterogeneity of the region of interest and therefore may provide similar outcomes to those obtained using entropy features from the COM.

CHAPTER SUMMARY: This chapter considered the use of an alternative technique for calculating texture within an MR image, using the max-min technique. The results validate those from the previous chapter, using the COM model, indicating that the grey-level pixel distribution within the MR images contains information that can be used for differentiating between cancer subtypes.

CHAPTER SIX: PROSPECTIVE PATIENT STUDY

OUTLINE OF CHAPTER: The previous chapters have demonstrated the usefulness of texture analysis in differentiation of malignant, benign and normal tissue. It has also demonstrated initial capability in categorising malignant tissue into histological and molecular subtypes with reasonable classification accuracies.
This chapter considers the use of these classifiers on a ‘blind’ test set in order to truly assess the usefulness of the technique in a clinical environment.

6.1 INTRODUCTION

It was demonstrated in Chapter Four that classification of lesions according to various categories of subtype using texture analysis alone could provide good classification accuracies and ROC areas. It was also demonstrated that there were significant differences in the raw feature values between a number of the different breast cancer subtypes.

Prior to concluding whether texture analysis can potentially be utilised as a clinical tool, it is essential to prospectively assess its ability to accurately classify data using the previously generated predictor model from the training set. By using data that the model has never ‘seen’ before, it is then possible to assess the likelihood of the technique becoming useful in a clinical setting.
This chapter considers a prospectively recruited patient lesion dataset in order to assess the utility of the model in predicting lesion subtype.

6.2 MATERIALS AND METHODS

6.2.1 Subjects

Women with the same inclusion criteria as outlined in Chapter 4.2.1, and scanned between June 2012 and August 2013 were prospectively recruited into this study. Again, only lesions greater than 8mm were included.

Histopathological core biopsy data was obtained after MRI data had been obtained, analysed and classified.

6.2.2 MR Imaging

All MR imaging was performed in an identical manner to that described in Chapter 4.2.2.

6.2.3 Texture Analysis and Classification

All texture analysis was performed as previously described in Chapter 4.2.3. All co-occurrence matrix features were considered, as well as entropy only features. Classification models created in Chapter 4 using Weka were applied to the new data and the classification of the data recorded. Pathology data was then compared to the predicted cancer subtype and an accuracy of the model calculated. ROC curves were then generated using Weka.

6.2.4 Classification of Entire Dataset

Due to the larger number of categories for the molecular subtypes, it was unclear whether poorer classifications could be attributed to inadequate models or due to an insufficiency of data in the training set to build an accurate predictive model. Therefore, once the pathology findings for the test data set were un-blinded, both the training and test datasets were combined in order to provide one single, large dataset. Classification of this data was then performed using Weka and an external validation method in which a percentage of the data is withheld in order to test the model, thus producing a training and test set within the data. The percentage split of this data was altered in order to assess whether classification accuracies could be further improved with larger training data sets.

6.2.5 Signal to Noise

Measures of signal to noise were also carried out in accordance with the method in Chapter 4.2.4.

6.3 RESULTS

6.3.1 Patient Cohort

Within our patient cohort that was prospectively recruited over a 15-month time period, there were a total of 75 lesions identified in 57 symptomatic women (age range: 27-85; mean age 56 years) that could be used in our study (over 8mm with pathology data obtained within one month of MR examination and consenting to images being used for research purposes). A summary of lesion characteristics is summarised in Table 6.33.

Of these lesions, 56 were identified on images that were acquired on the 1.5 T system and 19 were identified on 3.0 T examinations.

Patient Age	27-85 (median 55.5 years)
Maximum lesion diameter	12.5 mm (range: 8.2-24.5mm)
Histological Subtype	Ductal: 48; Lobular: 22; DCIS: 3
Histological Grade	Ductal- Gd2: 12; Gd3: 36 Lobular- Gd2: 18; Gd3: 4
Molecular Subtype	Luminal: 46; HER2: 11; TNBC: 15

Table 6.33- Summary of population characteristics of lesions included in prospective set.

6.3.2 Classification of Histological Subtype and Grade

Full histological data was only available for 73 of the 75 lesions included in the test set. The predicted histological subtype of these lesions is presented in Table 6.34.

There was excellent classification accuracy using the created model and all features with an accuracy of 72.5% and ROC area of 0.823. The classification results using the model created on the entropy features alone resulted in a poor classification accuracy with low ROC area, which is largely to be expected based on the poorer classification accuracy associated with the training set using these features alone.

			Predicted		*Classification Accuracy and ROC Area*
		Ductal	*Lobular*	*DCIS*	
Actual	Ductal	40 (35)	8 (13)	0 (1)	All: 72.5%; 0.823 Ent: 55.0%; 0.543
	Lobular	10 (17)	12 (5)	0 (0)	
	DCIS	1 (2)	1 (0)	1 (0)	

Table 6.34- Predicted histological subtype for lesions identified, as well as actual subtype as obtained from pathology reports and the classification accuracy and ROC areas for the created model. Figures in brackets are those predicted using entropy features only

The ROC curves generated for the histological classification of the data using the predictor model using all features, and entropy features alone are presented in Figure 6.48 (a) and (b) respectively.

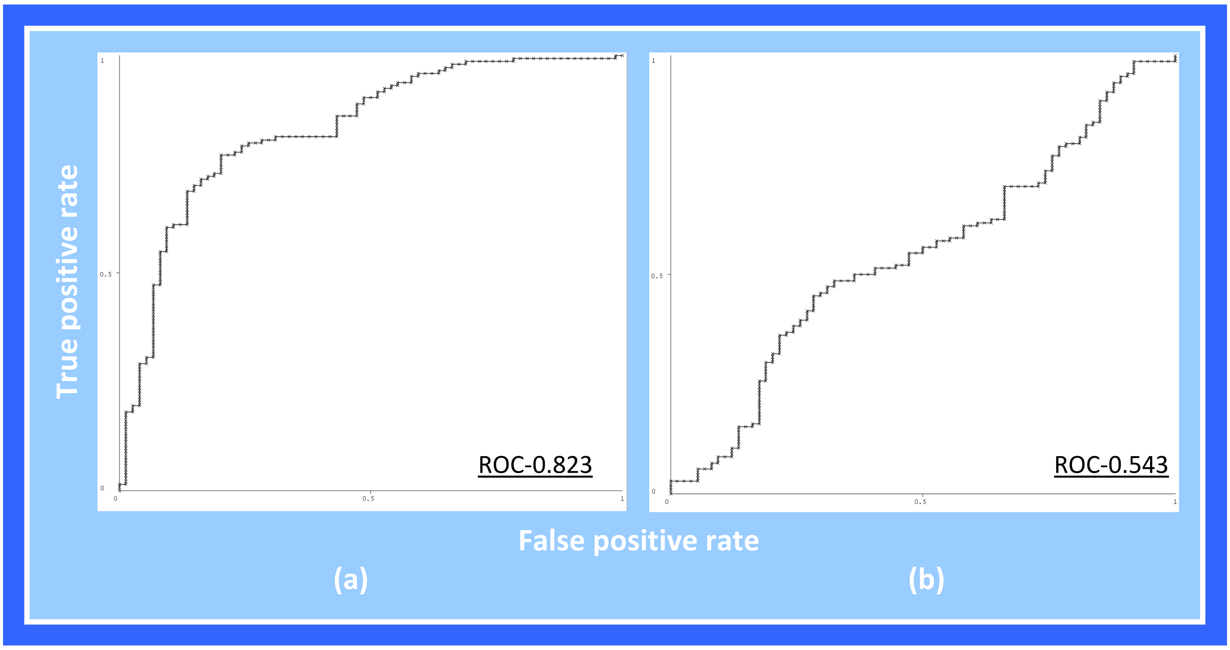


Figure 6.48- ROC curves for classification in terms of histological subtype using (a) all features and (b) entropy only features

After histological grade predictive classification had been carried out, the histology of lesions was identified in order to classify data into ductal and lobular subtypes prior to classification of histological grade. This step was performed in order to produce a simpler model and assess the performance on grade alone, rather than adding in another variable.

The histological grade predictions for both ductal and lobular cancers are provided below in Table 6.35 and the ROC curve obtained for classification of lobular cancer grade represented graphically in Figure 6.49.

		Predicted (all features)		*Classification Accuracy and ROC Area*
		Grade 2	*Grade 3*	
Actual	**Ductal**			
	Grade 2	7	7	All: 66.0%; 0.746
	Grade 3	8	26	Ent: 59.0%; 0.494
	Lobular			
	Grade 2	16	1	All: 87.9%; 0.944
	Grade 3	2	3	Ent: 78.8%; 0.699

Table 6.35- Predicted histological grade for ductal and lobular lesions identified, as well as true histological grading and calculated classification accuracies and ROC areas for the created model.

The classification accuracy for ductal cancers is good when all features are taken into consideration in the predictive model, however is poor for only entropy features. For lobular cancers, we had excellent classification accuracy, particularly when all features were incorporated in the predictive model, with almost 90% classification accuracy and ROC of 0.944. However it should be noted that there were only 4 lesions that were classified as Grade 3 lobular cancer and only around 20% of the lesions used in the training set on which the model was based were Grade 3 and therefore there remains the possibility that this model may not be optimal.

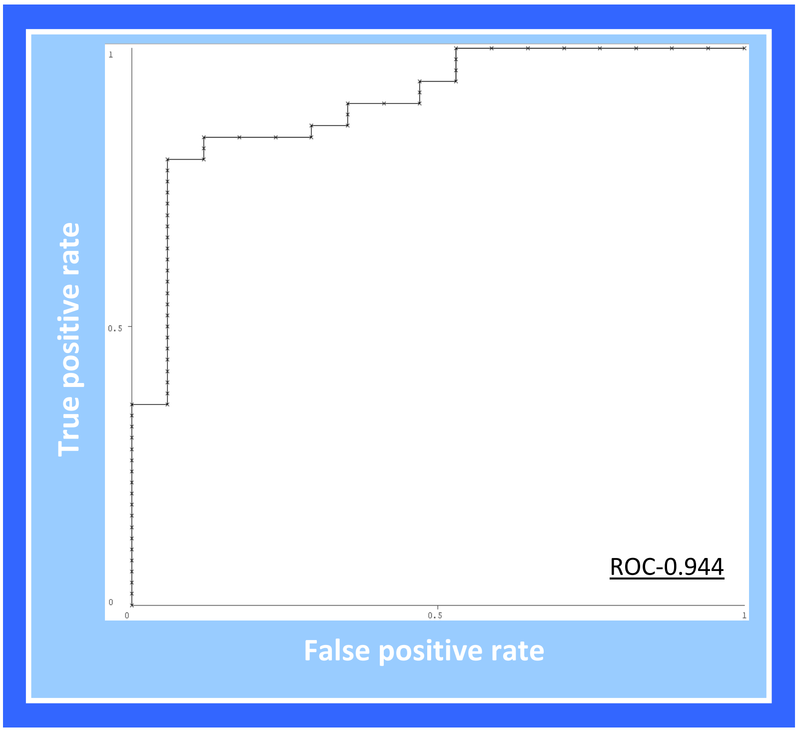


Figure 6.49- ROC curve for grade classification of lobular cancer using all features

6.3.3 Classification of Immunohistochemical Subtype

Immunohistochemical information was available for all 75 lesions in the cohort. All predictive receptor status data is presented in Table 6.36. Again the models based on the entire set of texture features perform much better that those models produced only on the entropy features. The triple negative vs. all other classifications results in the best data classification with over three quarters of the data correctly classified with a reasonable ROC area. For both the ER and HER2 status cancers, around two-thirds of data is classified correctly, although there are still high values of ROC area achieved.

		Predicted		*Classification Accuracy and ROC Area*
		Positive	*Negative*	
Actual	**ER Status**			
	Postive	42 (42)	8 (8)	All: 66.7%; 0.732
	Negative	17 (21)	8 (4)	Ent: 61.3%; 0.441
	HER2 Status			
	Positive	11 (13)	16 (14)	All: 68.8%; 0.762
	Negative	8 (15)	40 (33)	Ent: 61.6%; 0.598
	TN Status			
	TNBC	3 (1)	12 (14)	All: 76.8%; 0.726
	Other	5 (3)	55 (57)	Ent: 77.2%; 0.440

Table 6.36- Predicted immuniohistochemical subtype for ER, HER2 and Triple Negative receptor status. True subtype and calculated classification accuracies and ROC areas for the created model are also presented. Figures in brackets are those predicted using entropy features only.

6.3.4 Classification of Molecular Subtype

Molecular subtype information was again available for 72 lesions which were identified as invasive cancers and the actual subtype and predicted subtype values are as presented in Table 6.37. The molecular subtypes were classified into luminal, HER2 and triple negative cancers (TNBC). From this data, there is reasonable classification accuracy when all features are considered, but when only entropy features are considered, the classification results are disappointing.

		Predicted			*Classification Accuracy and ROC Area*
		Luminal	*HER2*	*TNBC*	
Actual	Luminal	42 (41)	1 (2)	4 (3)	All: 65.5%; 0.755 Ent: 58.6%; 0.437
	HER2	8 (9)	2 (1)	1 (1)	
	TNBC	10 (13)	0 (1)	4 (1)	

Table 6.37- Predicted molecular subtype. True molecular subtyping as identified using pathology, as well as calculated classification accuracies and ROC areas for the created model are presented. Figures in brackets are those predicted using entropy features only.

ROC curves generated from these predictive classifications are shown in Figure 6.50, where it is evident that the classifications based on all features result in significantly improved results.

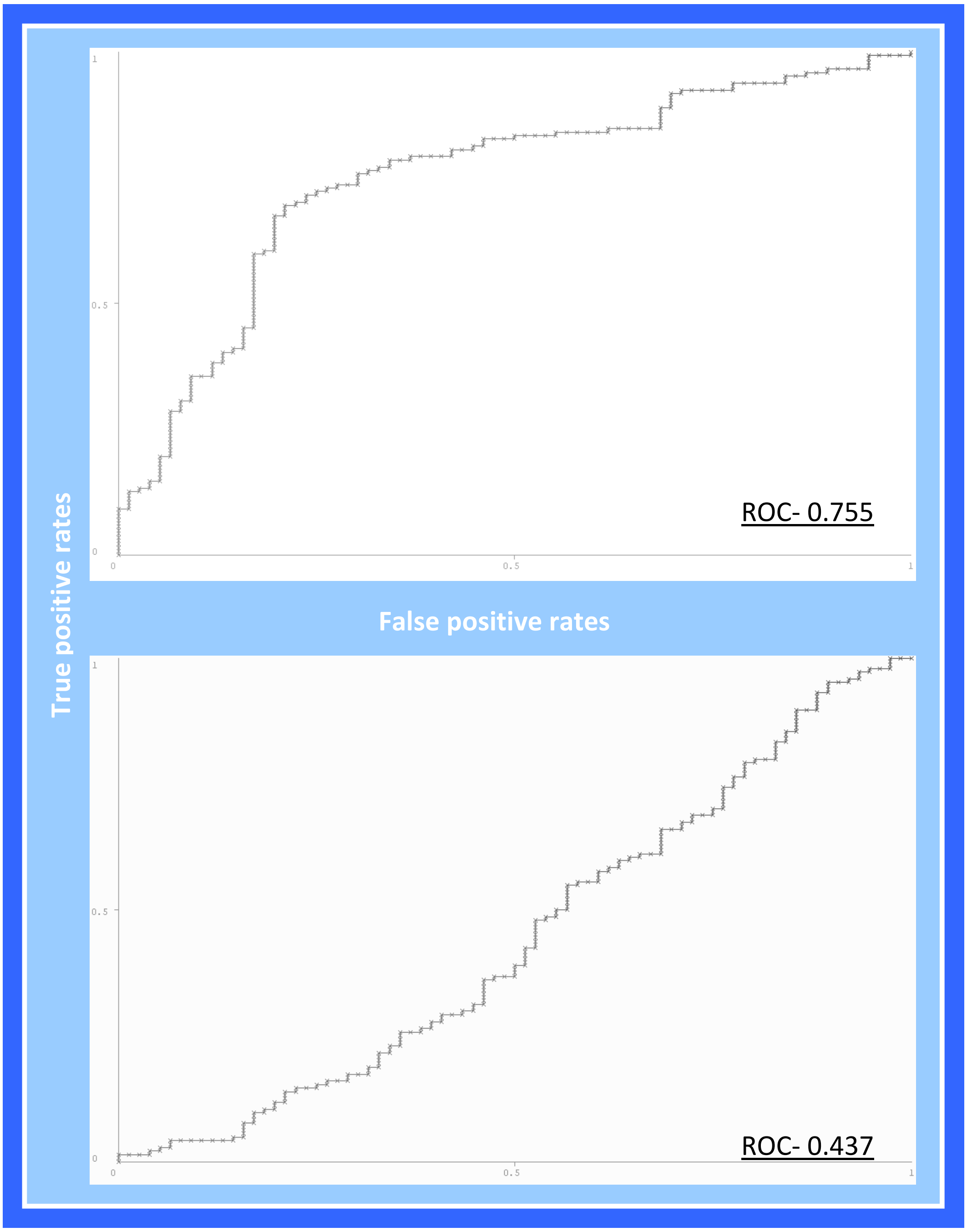


Figure 6.50- ROC curves for molecular subtypes classifying using all COM feature (top) and only entropy-based features (bottom)

6.3.5 Classification of Entire Dataset

Combining the training and test datasets resulted in a total of 212 lesions with full molecular subtyping information available. These were classified by setting different percentages of data to be used as training data with the remainder used to test predictive models. Data splits chosen for training data were 66% (as most representative of current numbers in each dataset), 70%, 75% and 80%. Classification accuracies and obtained ROC areas are shown in Table 6.38. The effect of changing this data split on classification into the three molecular subtype categories is summarised in the ROC plots in Figure 6.51.

Data Split	*Training Cases*	*Test Cases*	*Classification accuracy; ROC area*	
			All features	*Entropy features*
66%-34%	140	72	73.9%; 0.793	60.8%; 0.535
70%-30%	148	64	73.5%; 0.776	60.7%; 0.549
75%-25%	159	53	73.0%; 0.776	59.5%; 0.526
80%-20%	170	42	71.0%; 0.803	53%; 0.521

Table 6.38- Summary of classification results when training and test data considered together and a percentage of data withheld as test data. Classification results are presented as: [% accuracy; ROC area]

For the 66% data split, which results in numbers in the generated training and test set approximately equal to the numbers from the true training and test sets, classification accuracies are slightly improved (65.5% vs. 73.9%) although on a similar order of magnitude.

By increasing the number of cases that are included in the generated training dataset, the classification accuracies are relatively unchanged, however the ROC areas are improved, as is demonstrated in Figure 6.51, where there is a shift to the left of the plot, indicating improved classifier performance.

When 80% of the data is used as a generated training, the classification accuracy of the generated test data is good, with 70% of the data accurately classified and an ROC area of 0.803. There is less of an improvement to the models generated using only entropy features.

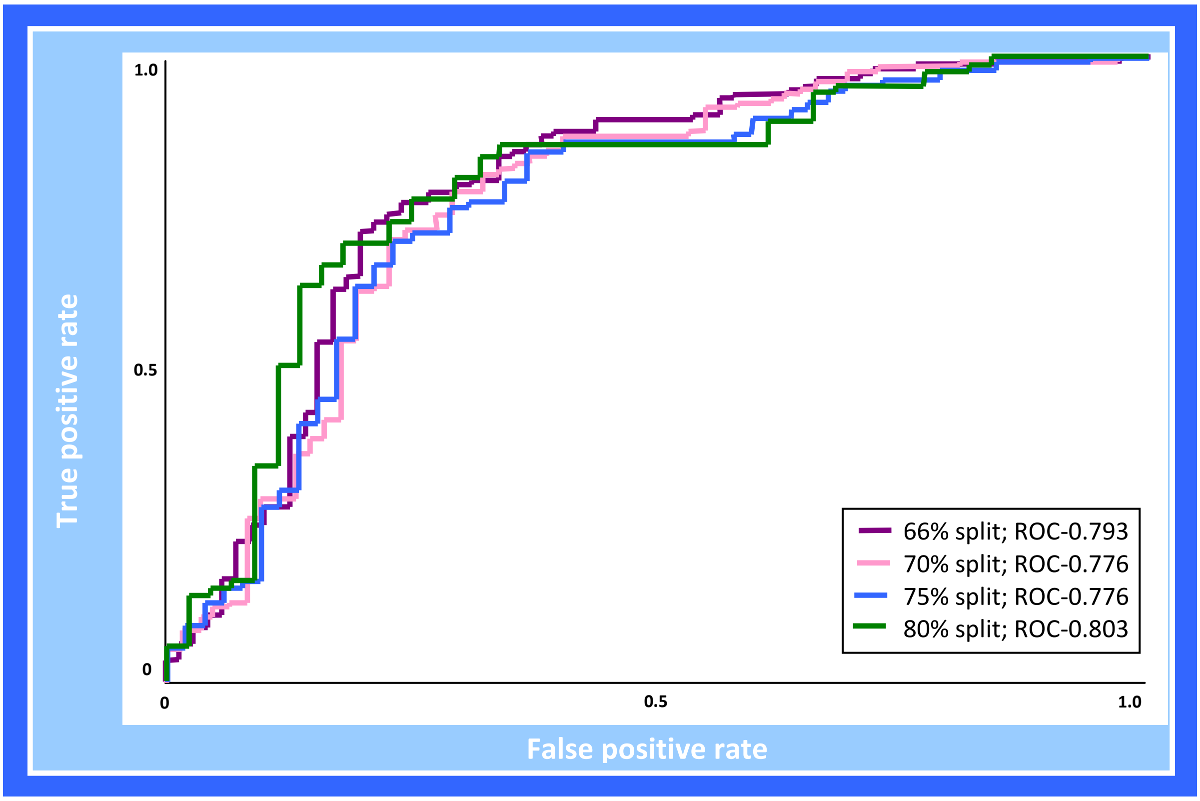


Figure 6.51- ROC curves for classification of molecular subtypes for various different splits of entire training and test data set

It should be noted that for the cases where a larger percentage of the data is used to generate a training set, this ultimately results in a lower number of cases in the test set, which could potentially influence the classification accuracies due to a smaller number of cases on which to test the model. For this reason, splits greater than 80%/20% were not considered. It seems possible, however, that an increase in lesion numbers could significantly improve classification accuracies.

6.3.6 Signal to Noise

Signal to noise measurements were calculated and compared as described previously and are presented in Figure 6.52. Error bars depict the standard deviation of the SNR measures in each subtype category. Again, large standard deviations are likely to be attributable to the distribution of scans performed on 1.5T and 3.0T.

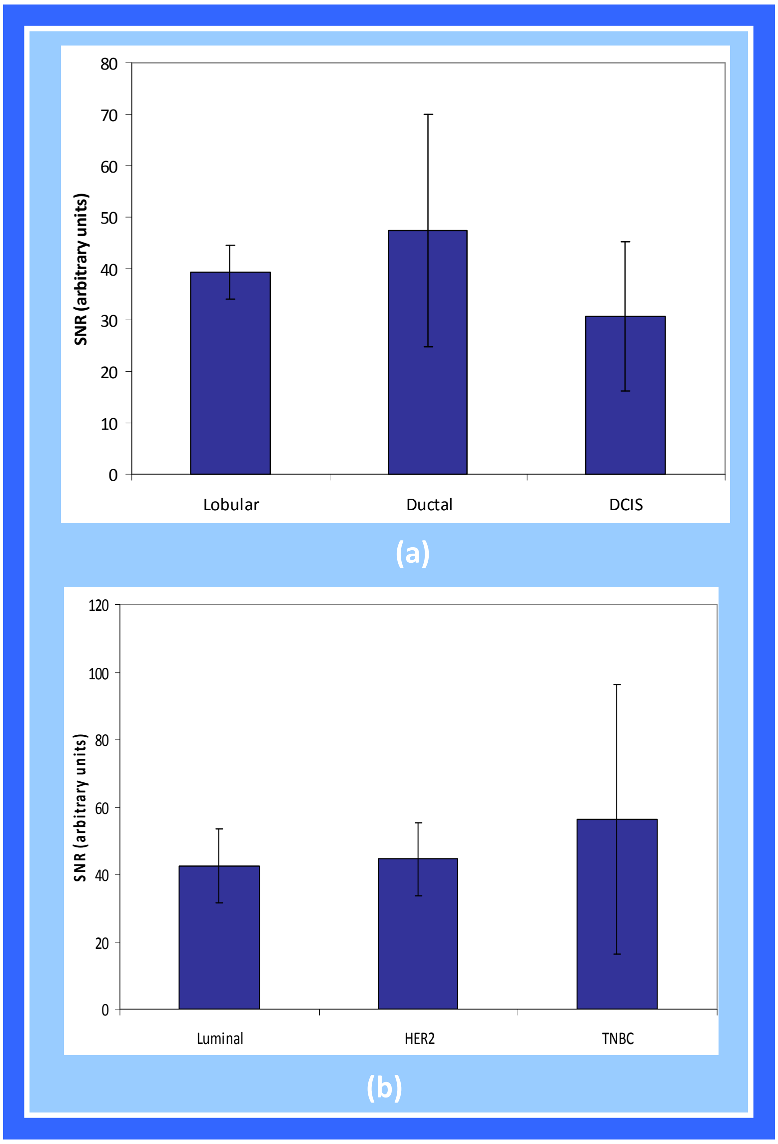


Figure 6.52- Summary of signal to noise measurements for (a) histological subtype groups and (b) molecular subtype groups

Statistical pair-wise comparisons were performed between each histological and molecular subtype using a students t-test, and one way ANOVA across all subtypes, and results of statistical analysis are presented in Table 4.28. There were found to be no significant differences in measured SNR values in each pair-comparison, suggesting that differences in texture analysis were unlikely to be attributable to this particular factor.

Histological Subtype Comparison	*SNR comparison p-value*	*Molecular Subtype Comparison*	*SNR comparison p-value*
Ductal vs. Lobular	p=0.206	Luminal vs. HER2	p=0.443
Ductal vs. DCIS	p=0.315	Luminal vs. TNBC	p=0.548
Lobular vs. DCIS	p=0.130	HER2 vs. TNBC	p=0.856
Histological subtype	p=0.204*	Molecular subtype	p=0.614*

Table 6.39- Summary of statistics performed on SNR differences between different histological and molecular subtypes of cancer. There were no significant differences ($p<0.05$; students t-test; *$p<0.05$; ANOVA)

6.4 DISCUSSION

Results presented in this chapter extend the work carried out in Chapter 4 to apply the classification models created on the previous training set to a blind test set which had not been used in the model creation. Using this technique, a true representation of how the models are likely to perform in a routine clinical environment is obtained and can therefore provide a measure of the usefulness of the technique as a clinical tool.

Classification of the test data set into ductal, lobular and DCIS histological subtype resulted in good classifications overall when all the COM features were used for model creation. Accuracy was over 70% with an ROC area of 0.823.

Histological grade also resulted in good classification accuracies across ductal and lobular cancers, although it should be noted that there were only a small number of Grade 3 lobular lesions.

When data was considered according to molecular subtype, classification accuracies were lower than anticipated with two thirds of the data well-classified, with only fair ROC values (0.755).

Once pathology data was available for all lesions, the training and test data sets were combined to provide one larger data set comprising 212 lesions with full molecular subtype information available. Training and test datasets were then generated from this data randomly using a percentage split method. For these, the classification accuracy was slightly improved compared with the true test data set predictions, however there is no control over the generation of the datasets and therefore it is unclear how many of each lesion type there is in each set, and also no method of knowing how many lesions were identified on 1.5T scans compared with 3.0T scans. This means that there could potentially be differences in SNR that we cannot quantify. While Weka can stratify the data when creating the training and test data sets to try and ensure a similar underlying population of data points in each group, it can do this only on one variable (e.g. lesion category) and there is then no control over further categorisations (such as scanner field strength). Unfortunately this provides a limitation to using this means of training/test set creation.

In Chapter 4, it was demonstrated that entropy features appeared to be statistically different between the HER2 and luminal cancers, suggesting that these features would result in a good classification accuracy, however in all predictive classifications performed, the entropy features result in disappointing classification accuracies, with meaningless ROC values around 0.5.

However, as we obtained good training classification accuracies, and statistically significant differences in feature values between a number of the molecular and histological subtypes, it is possible that the existing study is underpowered to create a robust training model, particularly when this model is created using only a few of the calculated features as in the case of the entropy features. However, there are promising results when all COM features are used, and when the data is combined and a training and test set generated using a larger number of data points and therefore this technique warrants further investigation by increasing the number of cases significantly.

The major limitation in this study was the number of patients, and hence cancers, that were included. However, this was due to the inclusion criteria and matching as closely as possible the lesions that were included in the training data set to ensure there was no influence of lesion or ROI size on the outcome of the study. A basic, retrospective power calculation (G*Power, v3.1.5; Kiel University, Germany) [158] performed using the mean values for the training dataset in Chapter 4 suggests that in order to classify differences between luminal, HER2 and TNBC lesions with a 95% power, using entropy features alone, a dataset of 250 lesions would be required (split with approximately 60% of the data being luminal cancers). This suggests that our study must be extended considerably in order to fulfil this criterion.

Again, we had a low number of DCIS cases, which reflects the referral criteria for breast MRI examinations and was a factor that was out with our control.

Predicted grade was performed after pathology results were available to allow us to split the data into ductal and lobular cancers prior to classifications into grade. Whilst a full model could have been created in order to differentiate simultaneously between ductal and lobular cancers as well as grade, this step was not performed as this would also have required DCIS data to be included, and we were also trying to identify the usefulness of TA in each individual classification of cancer- for which we separated histological sub typing and grade classification.

Classification of lesions into their respective molecular subtypes is still an evolving technique and in a recent study performed by Mackay et al. it was found that even in cases where expert microarray profilers were asked to classify lesions, the inter-observer agreement is reproducible (kappa>0.81) only in the classification of basal like and HER2 lesions [159]. For classification between luminal cancers the kappa value was less than 0.61 [159].

Therefore, an image based system that can reliably differentiate between luminal, HER2 and TNBC with an ROC area of 0.733 and classification accuracy of 64.3% could potentially provide added value for decision making in a clinical environment, prior to biopsy information being available.

The data provided in this chapter demonstrates that while we have insufficient data points to demonstrate with certainty that texture analysis can be used prospectively in classification of lesion subtype in a clinical environment, there is initial promise. Further patient recruitment and larger data sets are required in order to increase the training and test data sets in order to conclusively demonstrate whether the technique can produce accurate subtype predictions, but initial results suggest that this is a real possibility and one that warrants further investigation.

CHAPTER SUMMARY: This chapter considered the usefulness of texture analysis predictive models as created previously applied to a blind test set. Initial results suggest that using all features, classifications of histological subtype and grade can be performed with excellent results.

The results from the molecular subtype classifications are inconclusive- which is most likely to be due to insufficient cases in order to build a robust model. However, there is initial promise shown and a brief investigation on the influence of training dataset size suggests that increasing the numbers could result in improved data classification.

CHAPTER SEVEN: TEXTURE ANALYSIS IN TREATMENT RESPONSE

OUTLINE OF CHAPTER: Previous chapters have outlined the usefulness of TA in the diagnostic setting for classification of malignant and benign lesions as well as histological and molecular subtypes. This chapter considers the use of texture analysis in early response to neoadjuvant chemoteherapeutic treatment and whether ultimate outcomes can be predicted based on texture analysis measures.

7.1 INTRODUCTION

Neoadjuvant chemotherapy (NAC) is drug treatment that is delivered prior to breast surgery with the aim of reducing the size of a locally advanced breast cancer into an operable tumour that may previously have been inoperable, or to reduce the extent of surgery required [160]. However, such treatment regimes are not effective in all patients and some women will undergo intensive therapy with potentially distressing side effects, and the disease will remain resistant or in some cases will have continued to grow throughout the treatment. Magnetic resonance imaging is becoming more commonly used in the monitoring of treatment due to the lack of ionising radiation, allowing repeated examinations and dynamic contrast enhanced imaging provides a measure of tumour vascularity. However, it has been reported that MR can under- or over- estimate residual tumour in around 30% of cases [161] and therefore more quantitative measures and earlier response markers are required in order to increase the accuracy of such response measures and allow them to be implemented in routine practice.

This chapter considers the use of texture analysis in a cohort of patients undergoing neoadjuvant chemotherapy at our institution and correlates this with traditional response evaluation criteria- using radiological assessment and final resectional pathology data.

7.2 MATERIALS AND METHODS

7.2.1 Subjects

All women with biopsy-proven cancer who were scheduled for neoadjuvant chemotherapy (NAC) between January and August 2013 were included in this study. Women were scanned after diagnosis but prior to commencement of chemotherapeutic treatment. Only patients who had consented for their examination to be used for research and developmental purposes were included, and no active recruitment process took place. Findings from this study bore no impact on the ultimate course of the patient treatment or management.

As was previously described in earlier chapters, only lesions greater than 8mm were included in the study. As NAC is rarely offered for cancers less than 2cm in size all patients who consented were included, and image analysis was performed blinded to the treatment outcome for each patient- this information was obtained after texture analysis had been performed and prior to statistical evaluation of the data.

7.2.2 MR Imaging

All MR examinations were performed on the 3.0 T MRI Scanner (Trio; Siemens, Erlangen). As per the standard NAC protocol, patients were scanned prior to starting their chemotherapeutic treatment, at interim (either post cycle 2 or cycle 3) to assess early response and at the end of their treatment in order to assess final response.

All acquisition conditions were as described previously for the 3.0T protocol (Section 4.2.2) and texture analysis was performed on the baseline and interim examinations in order to identify if TA could identify or characterise early response.

7.2.3 Texture Analysis

Images from baseline and interim examinations were considered side-by-side in order to match slices depicting lesions from each examination to best match the region of analysis between the two separate examinations.

Texture analysis was then performed as previously described, by placing 10×10 pixel ROIs on the three slices with largest lesion dimensions. There was also a 10×10 ROI placed in healthy, normal contralateral breast tissue (see Figure 7.53). In order to ensure that changes measured were real and not due to changes related to scan-scan conditions, average standard deviation of the background noise was measured by placing ROIs as shown in Figure 4.28 for all cases. Comparisons of baseline and interim examinations were then made using paired t-tests, with $p<0.05$ considered significant. SNR was not felt to be a valid measure in this instance as this would reflect contrast uptake kinetics within the lesion, adding a confounding factor to interpretation due to the link with response.

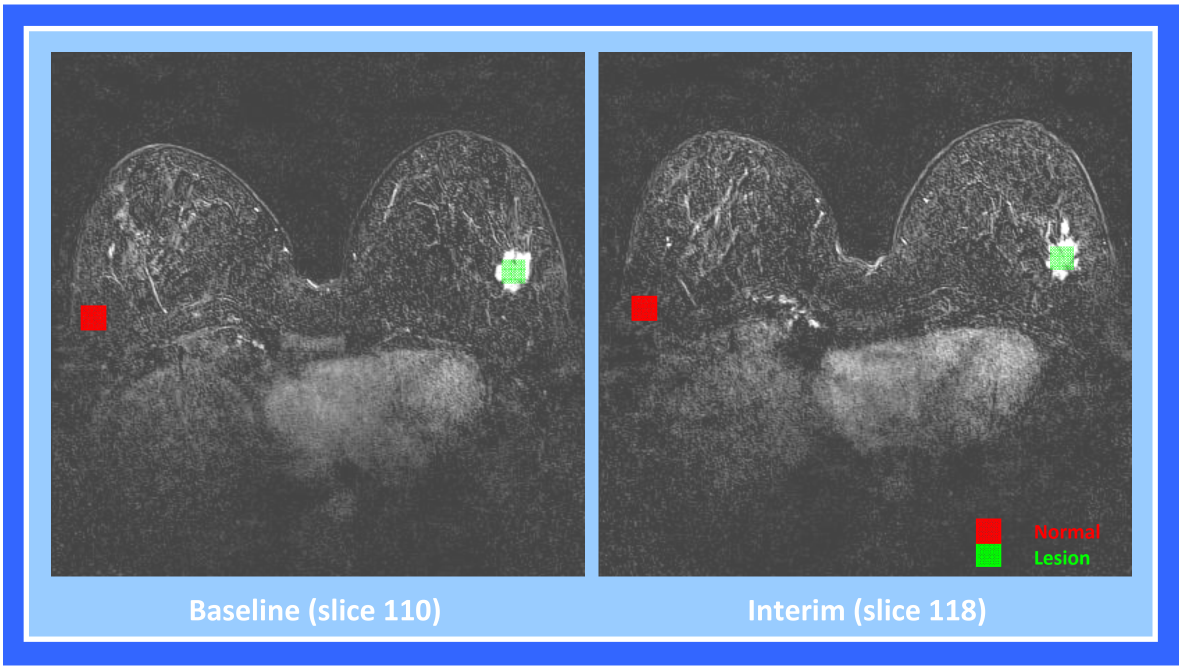


Figure 7.53- Analysis protocol for comparing texture analysis at baseline and interim MRI examinations. Regions of interest were placed in normal and lesion tissue

Comparison of texture features in respect to response was considered in a number of ways:

1. Comparing the number of significantly different features between baseline and interim examinations for each of the response categories, for normal tissue and lesion regions.

2. Comparing baseline and interim absolute features using a Wilcoxon test for each response category.
3. Subtracting baseline and interim features for each response category.
4. Considering baseline values alone with respect to response categories to identify any features that may indicate whether lesions will respond or not.

7.2.4 Assessment of Response

The final response to chemotherapy was assessed by an expert breast MRI radiologist using RECIST criteria (Response Evaluation Criteria in Solid Tumours) [162]. This method defines response into four categories as outlined in Table 7.40. All size changes are measured relative to pre-treatment baseline values and are assessed at interim and final examinations. All RECIST classifications were obtained from the NHS radiology report from the MRI examinations. Progressive and stable diseases were considered as one category, as were complete response and minimal residual disease based on evidence from survival studies.

Response Category		*Criteria*
Complete Response	CR	disappearance of lesions
Minimal Residual Disease*	MRD*	Minimal enhancement, almost complete disappearance of lesions*
Partial Response	PR	>30% decrease in longest diameter sum
Progressive Disease	PD	>20% increase in longest diameter sum
Stable Disease	SD	small changes that do not meet other criteria

Table 7.40- RECIST criteria for response classification to treatment [162] (*note: MRD is not an official RECIST criteria but is used routinely at this institution)

As well as classification of disease using an image-based technique alone, the residual cancer burden (RCB) score was also calculated from final pathology [163]. This measure provides an assessment of tumour dimensions, cellularity and axillary node burden in the tumour after therapy and has been linked with disease-free survival [163] and therefore could potentially be a more useful prognostic marker with which to correlate texture results. RCB scores were retrospectively calculated as requested by an expert specialist breast pathologist (Dr Colin A. Purdie) and reported in terms of the RCB index as shown in Table 7.41 [163].

RCB Category		*% population with relapse at 5 years*
pCR	Pathological complete response	0.0 %
RCB-I	Minimal residual disease	21.6 %
RCB-II	Moderate residual disease	36.2 %
RCB-III	Extensive residual disease	52.2 %

Table 7.41- Residual cancer burden categories and appropriate survival statistics at 5 years post treatment

7.3 RESULTS

7.3.1 Patient Cohort

A total of 61 patients were identified for this study, however due to scheduling problems, 7 received their first cycle of treatment prior to MRI examination, 3 received their interim examination after four cycles of treatment and 2 patients had their treatment stopped after 4 cycles and therefore there was no true end of treatment assessment of response possible.

In the remaining 49 patients (age range: 30-79; mean age 50 years) included in the study, there were a total of 57 lesions identified over 8mm. All patients were scanned at baseline and then after cycle 2 (22 patients) or cycle 3 (27 patients) of their treatment, depending on scanner availability.

All patients received 6 cycles of treatment as shown in Table 7.42. FEC (fluorouracil, epirubicin and cyclophosphamide) is the most commonly used treatment regimen used at our institution for breast cancer patients and may be used alone or in combination with Docetaxyl (DOC) and/ or trastuzumab Herceptin (TRA). There were also four patients in this study who received a docetaxyl in combination with TDM1 (trastuzumab emtansine) for treatment of aggressive HER2 positive cancer. Trastuzumab and TDM1 treatments were only used where the breast cancer was HER2 positive.

First 3 cycles	*Second 3 cycles*	*No of patients*
FEC	FEC	26
FEC	DOC	8
FEC	DOC+TRA	11
TDM1+DOC	TDM1+DOC	4

Table 7.42- Treatment regimes for patients included in this study (see text for regimen descriptors)

7.3.2 RECIST Response Criteria

Of the patients that underwent treatment, 11 were considered to have had a complete MR response or have minimal residual disease (MRD), 27 were categorised as having a partial response (PR) and 11 were classified as stable or progressive disease (SD).

The number of significantly different features between baseline and interim examinations for each of the response categories was calculated using a Wilcoxon test with $p<0.05$ considered significant. The statistics were calculated for normal tissue and within the lesion and results are presented in Figure 7.54.

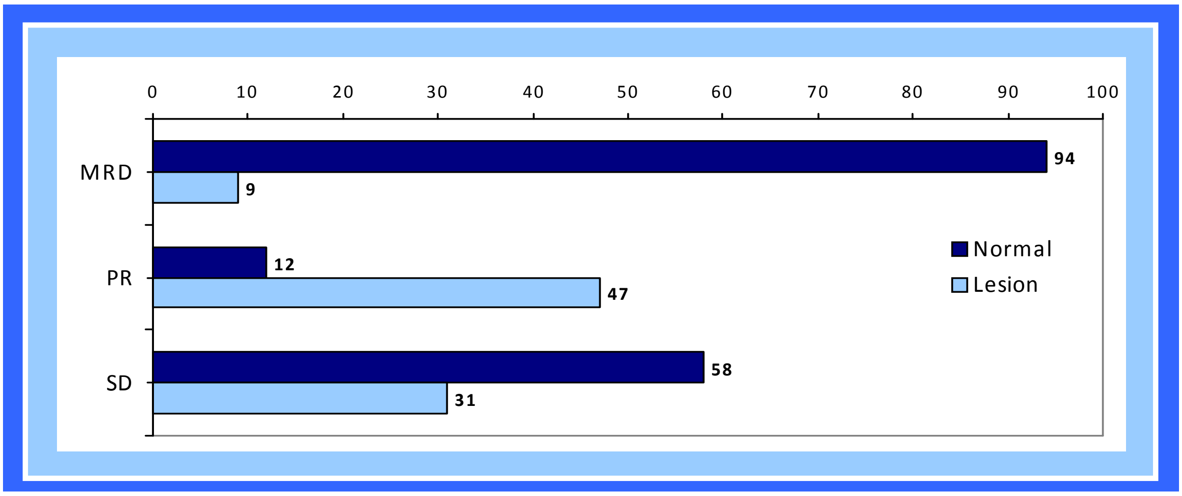


Figure 7.54- Number of significantly different features between baseline and interim examination for normal and lesion regions of interest, out of 220 calculated features (Mann Whitney U, $p<0.05$)

Statistical comparison of features was then performed between baseline and interim examinations for normal and lesion ROIs in each clinical response category as presented below in Table 7.43. Unfortunately, it appears that there are significant differences between the baseline and interim values for normal tissue in most of these categories and therefore it is unlikely that differences in lesion features between MRI examinations can be conclusively linked with the ultimate outcome of treatment.

	Normal			*Lesion*		
	MRD	PR	SD	MRD	PR	SD
AngScMom	<0.001	<0.001	<0.001	<0.001	0.013	0.117
Contrast	0.009	0.015	0.424	0.119	0.345	0.117
Correlat	0.785	0.763	0.976	0.097	0.566	0.757
DifEnt	<0.001	0.219	<0.001	<0.001	0.009	0.642
DifVarnc	0.018	<0.001	0.023	0.391	0.357	0.001
Entropy	<0.001	<0.001	<0.001	<0.001	0.067	0.221
InvDfMom	<0.001	<0.001	<0.001	<0.001	0.229	0.919
SumAverg	<0.001	<0.001	0.845	0.921	0.001	0.004
SumEnt	<0.001	0.009	<0.001	<0.001	0.252	0.547
SumofSqs	<0.001	<0.001	0.037	0.940	<0.001	<0.001
SumVarnc	0.010	0.079	0.335	0.078	0.054	0.325

Table 7.43- Statistical comparison between baseline and interim examinations for each texture feature for normal and lesion regions of interest in each clinical response category. Statistically different p-values are highlighted in grey (Wilcoxon test; $p<0.05$)

Considering the background values, there was no significant differences found in the background standard deviation value between baseline and interim examinations for any of the response categories ($p>0.281$, students paired t-test) suggesting measured differences were not attributable to scan-scan conditions.

The difference between baseline and interim features were calculated as shown in Figure 7.55 for the entropy features for normal and lesion regions of interest.

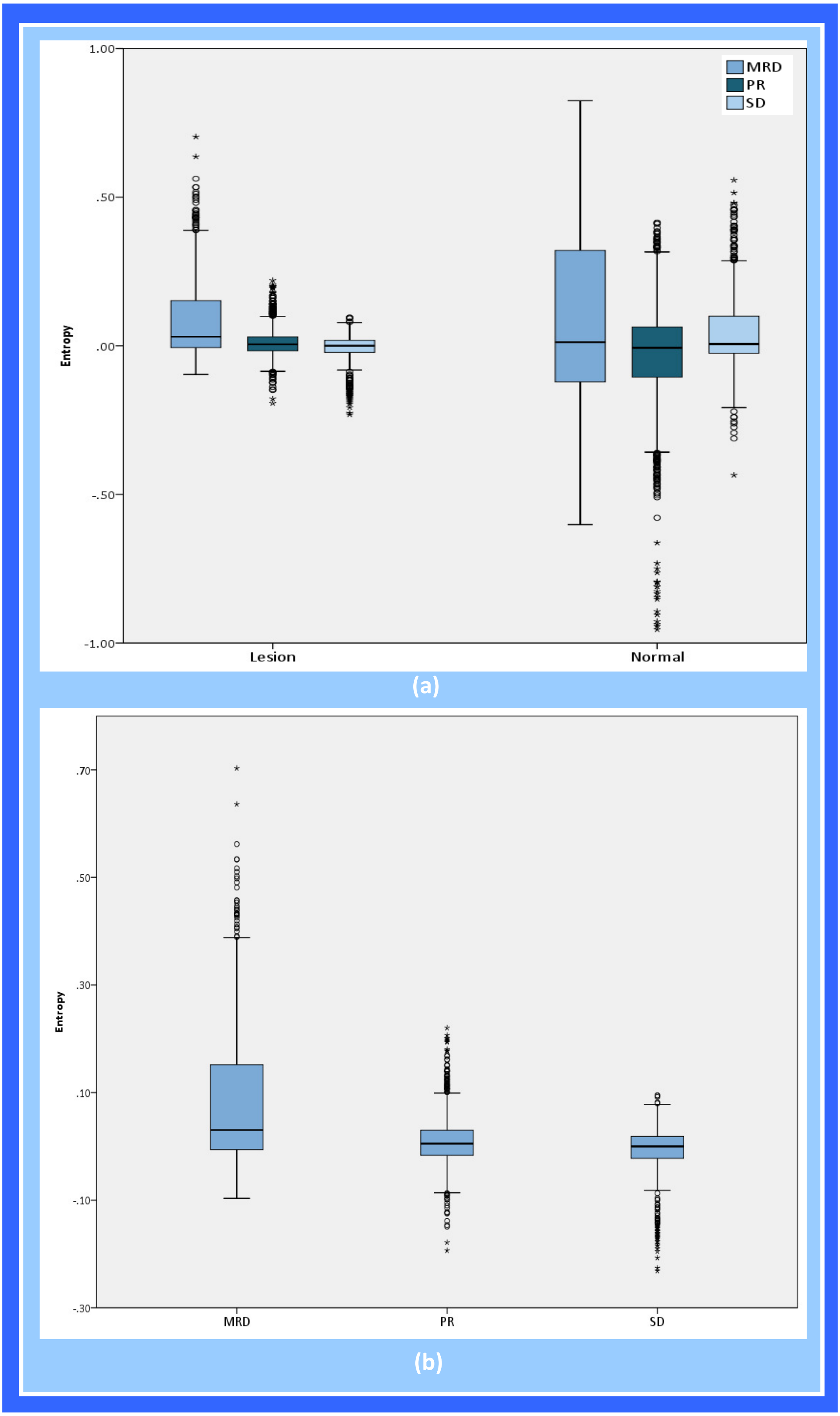


Figure 7.55- Difference in entropy features between baseline and interim examinations for normal and lesion regions of interest (a) and lesion only (b) in each clinical response category

The normal tissue shows a wide range of values, and therefore lesion entropy values only, which show a tighter range of values, are also presented separately in Figure 7.55(b). While statistically significant differences were obtained for a number of features (entropy, difference entropy, sum entropy and second angular momentum) between each response category, significant differences were also found in normal tissue (Mann-Whitney U; $p<0.05$) and therefore it was concluded that there was no real correlation between changes in texture values and ultimate clinical response to chemotherapy treatment as measured using RECIST criteria in these patients.

Finally, assessment of whether absolute texture features at baseline correlated with ultimate clinical response assessed using RECIST was performed. Statistical comparison of each texture feature between each clinical response category was carried out using a Mann-Whitney U test with a significance level of 0.05.

		MRD v PR	*MRD v SD*	*PR v SD*
Normal	AngScMom	0.567	0.015	0.036
	DifEnt	0.378	0.142	0.386
	Entropy	0.437	0.002	0.003
	SumEnt	0.473	0.014	0.022
Lesion	AngScMom	0.173	0.111	0.659
	DifEnt	0.683	0.811	0.502
	Entropy	0.199	0.156	0.719
	SumEnt	0.046	0.182	0.781

Table 7.44- Statistical comparison of baseline features between each clinical response category, significant values are highlighted in grey (Mann-Whitney U; $p<0.05$)

Results are presented in Table 7.44 for the four features which produce the most significant results. It is clear that there appears to be no correlation of baseline feature value with differentiation between response categories.

Graphical representation of these results is presented in Figure 7.56 for entropy and difference entropy features.

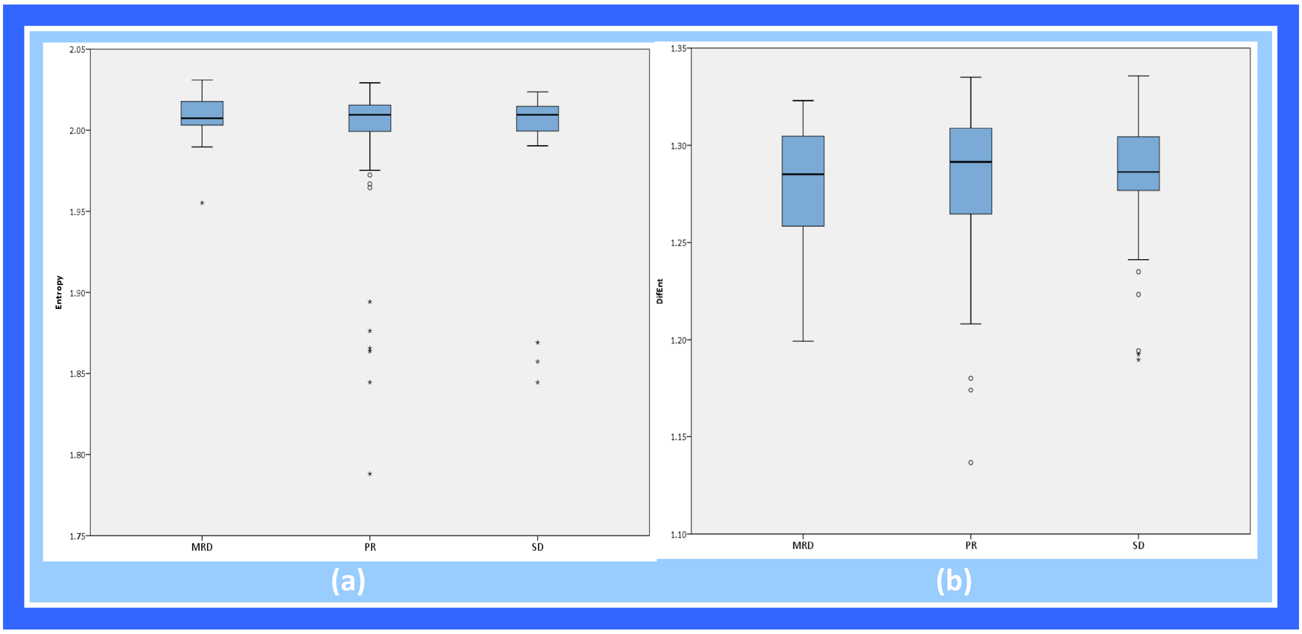


Figure 7.56- Baseline only values for each clinical response category for lesion regions of interest. Box-whisker plots are displayed for both entropy (a) and difference entropy (b) features

7.3.3 RCB Response Criteria

RCB scores were available for only 40 of the 49 patients included in the study and were categorised as shown in Figure 7.57.

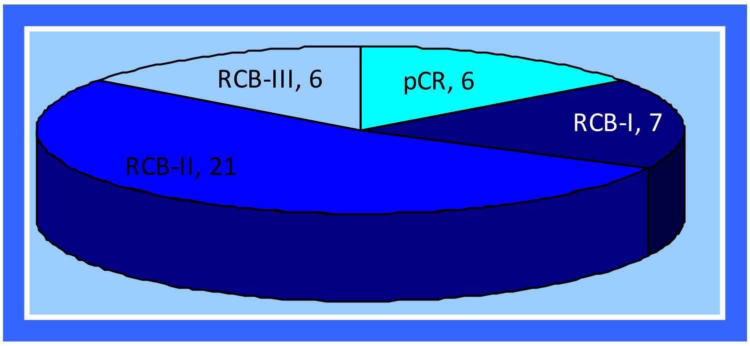


Figure 7.57- Breakdown of patients in each residual burden category

The number of significantly different features between baseline and interim examinations for each of the response categories was calculated using a Wilcoxon test with $p<0.05$ considered significant. The statistics were calculated for normal tissue and within the lesion and results are presented in Figure 7.58.

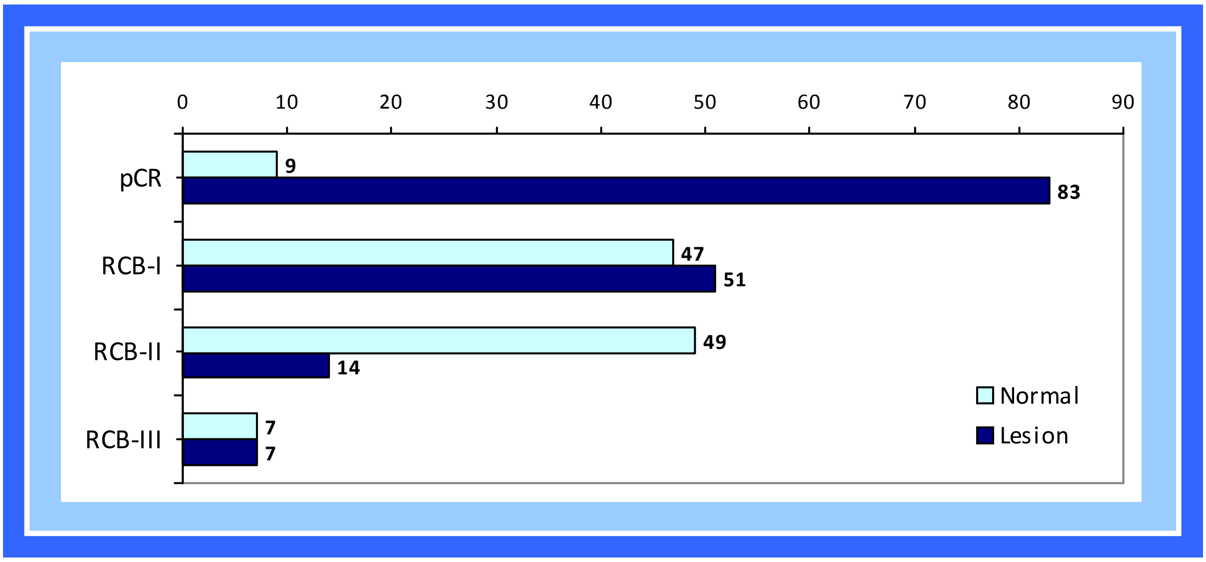


Figure 7.58- Number of significantly different features between baseline and interim examination for normal and lesion regions of interest in each residual cancer burden index, out of 220 calculated features (Wilcoxon test, p<0.05)

The texture features calculated at baseline and interim examinations for normal and lesion ROIs were then compared using a Wilcoxon test. This data is presented in Table 7.45 for each RCB category.

	pCR		*RCB-I*		*RCB-II*		*RCB-III*	
	N	L	N	L	N	L	N	L
AngScMom	0.791	<0.001	0.389	0.044	0.038	0.060	0.606	0.192
Contrast	0.171	0.161	0.293	0.017	0.920	0.691	0.226	0.888
Correlat	0.203	0.064	0.791	0.005	0.694	0.836	0.462	0.521
DifEnt	0.542	0.323	0.214	<0.001	0.046	0.809	0.406	0.542
DifVarnc	0.239	0.001	0.406	0.019	0.079	0.772	0.040	0.406
Entropy	0.767	0.001	0.355	0.047	0.046	0.029	0.673	0.085
InvDfMom	0.767	0.013	0.389	0.059	0.050	0.316	0.673	0.521
SumAverg	0.226	0.004	0.521	0.815	0.972	0.225	0.355	0.839
SumEnt	0.815	0.001	0.542	0.864	0.026	0.059	0.265	0.135
SumofSqs	0.406	0.913	0.040	0.563	0.940	0.727	0.767	0.339
SumVarnc	0.767	0.323	0.013	0.111	0.698	0.509	0.719	0.389

Table 7.45- Statistical results from comparisons of each texture feature between baseline and interim examinations in each residual cancer burden category. Significant p-values are highlighted in grey [N-normal, L-lesion] (Wilcoxon test; p<0.05)

From this data, the difference in lesion texture features between baseline and interim examination in the pathological complete response category (pCR) and minimal residual disease category (RCB-I) shows significant differences for angular second moment, entropy and difference variance features. For difference variance, angular second moment and entropy, this is coupled with no measured differences in normal tissue for these texture features. For the RCB-II category (moderate residual disease) there is also a statistically significant difference between baseline and interim measures of entropy features, although significant differences are also measured in normal tissue and therefore the meaning of this result is unclear.

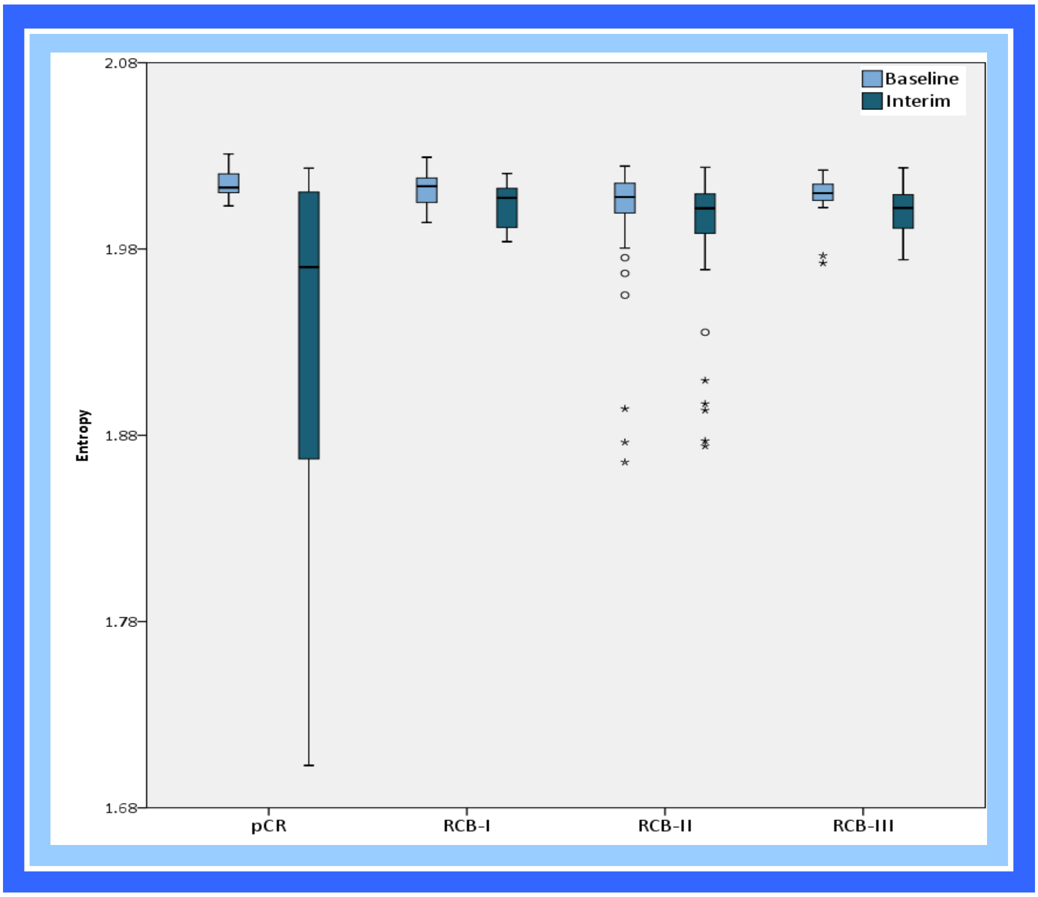


Figure 7.59- Absolute entropy values at baseline and interim for each residual cancer burden index category

The values for this entropy feature as measured in the lesion in each RCB category are shown in the box-whisker plot (Figure 7.59). All lesions show a general trend towards a reduction in the entropy parameter at the interim examination compared with baseline suggesting a reduction in heterogeneity within the lesion. The complete response category demonstrated a very wide range in the entropy values measured at the interim examination compared with all others.

Measures of the standard deviation of background noise between baseline and interim examinations were not significant for any of the categories: pCR: p=0.330; RCB-I: p=0.662; RCB-II: p=0.553; RCB-III: p=0.589 (paired student's t-test) therefore there were no differences attributed to scan-to-scan variations between categories.

The difference in feature values between baseline and interim examinations was calculated for each feature and each response category, considering the normal and lesion regions of interest. These are represented graphically in Figure 7.60 for the entropy features which resulted in most discriminatory parameters, and the lesion only regions of interest subtractions are also shown alone in Figure 7.60(b).

Statistical comparisons between the differences in texture features using Mann-Whitney U tests are presented in Table 7.46 for the four features demonstrating the most significance.

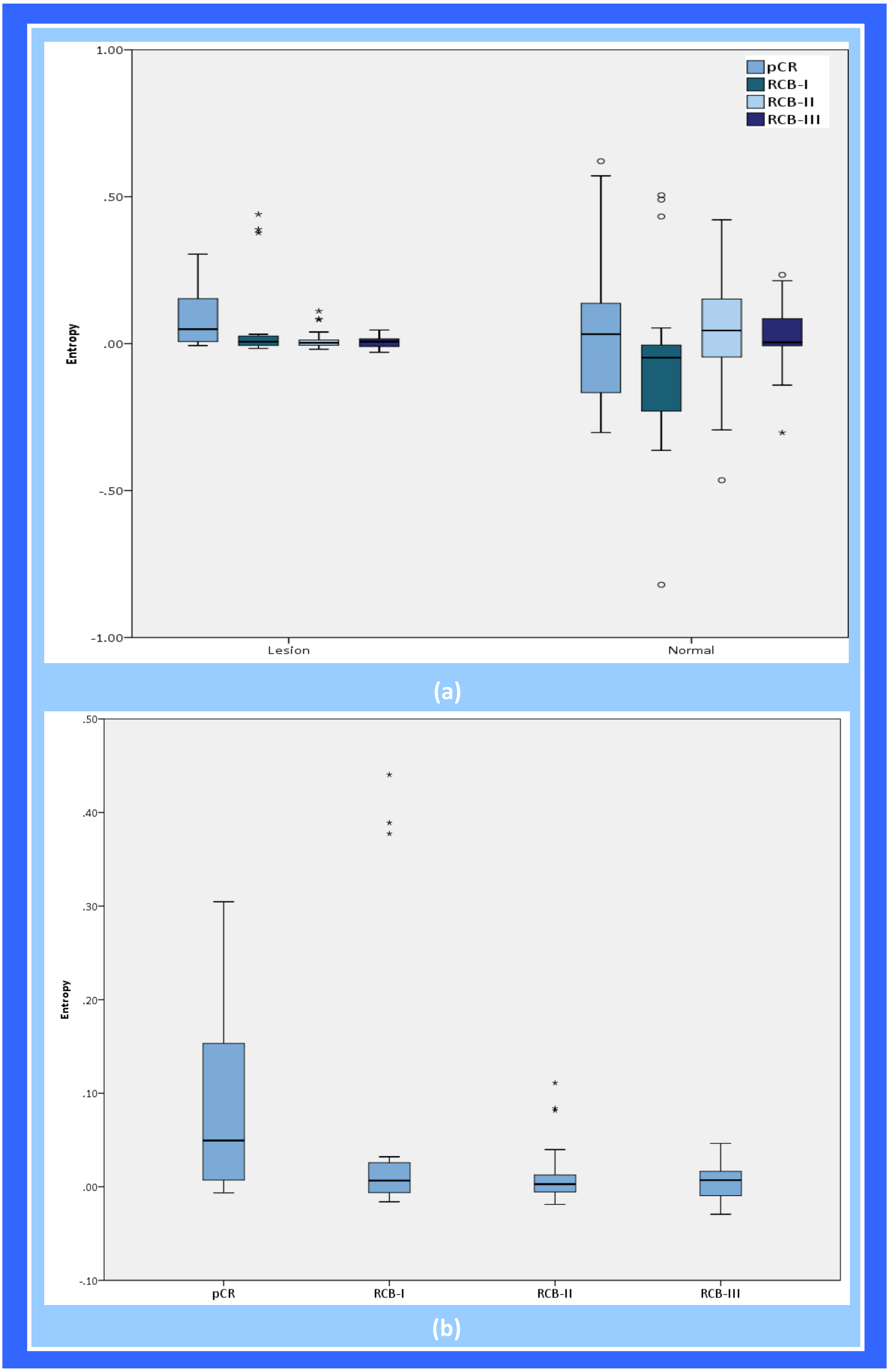


Figure 7.60- Difference in entropy features between baseline and interim examinations for normal, background and lesion regions of interest (a) and lesion only (b) in each residual cancer burden index category

From these results, the difference in texture features between baseline and interim examination as measured in normal tissue is not statistically significant. However, for pCR compared with both RCB-II & III and for RCB-I compared with RCB-II angular second moment, entropy and sum entropy appeared to consistently result in a significant difference when the change in texture features were compared between response categories (Table 7.46).

		pCR v RCB-I	*pCR v RCB-II*	*pCR v RCB-III*	*RCB-I v RCB-II*	*RCB-I v RCB-III*	*RCB-II v RCB-III*
Normal	AngScMom	0.213	0.561	0.300	0.084	0.444	0.381
	DifEnt	0.189	0.530	0.279	0.081	0.432	0.381
	Entropy	0.244	0.859	0.612	0.056	0.318	0.521
	SumEnt	0.259	0.791	0.526	0.076	0.394	0.464
Lesion	AngScMom	0.945	<0.001	0.006	0.003	0.083	0.435
	DifEnt	0.044	0.069	0.256	<0.001	0.008	0.864
	Entropy	0.708	<0.001	0.002	0.003	0.091	0.781
	SumEnt	0.485	<0.001	0.023	0.037	0.336	0.532

Table 7.46- Statistical comparisons for each region of interest in each residual cancer burden category. Significant differences are highlighted in grey (Mann-Whitney U; $p<0.05$)

From Figure 7.60 we can see that the difference in the entropy feature is higher for pCR than that of the other response categories, with a wider range of values compared with the other categories.

The differences measured in other response categories fell within a much tighter range of entropy values as shown in Figure 7.60, even for the RCB-II category which contained the most patients and therefore this is unlikely to be a population effect.

Finally, an assessment was made as to whether absolute texture features as measured on baseline examinations could be correlated with the final residual cancer burden as categorised in Table 7.41. Comparisons between response categories were made using a Mann-Whitney U test, and results are summarised in Table 7.47 for the four features which appear to provide most significant differences.

		pCR v RCB-I	*pCR v RCB-II*	*pCR v RCB-III*	*RCB-I v RCB-II*	*RCB-I v RCB-III*	*RCB-II v RCB-III*
Normal	AngScMom	0.155	0.439	0.331	0.079	0.084	0.117
	DifEnt	0.026	0.595	0.416	0.003	0.010	0.198
	Entropy	0.203	0.435	0.140	0.215	0.035	0.015
	SumEnt	0.114	0.541	0.189	0.066	0.026	0.047
Lesion	AngScMom	0.262	0.038	0.023	0.099	0.181	0.359
	DifEnt	0.105	0.050	0.013	<0.001	<0.001	0.235
	Entropy	0.443	0.045	0.046	0.085	0.179	0.319
	SumEnt	0.005	0.026	0.073	0.468	0.763	0.698

Table 7.47- Statistical comparison of baseline features between each residual cancer burden category, significant values are highlighted in grey (Mann-Whitney U; p<0.05)

Absolute values for these features in each response category are presented in Table 7.48 and are presented graphically for the entropy and difference entropy features in Figure 7.61.

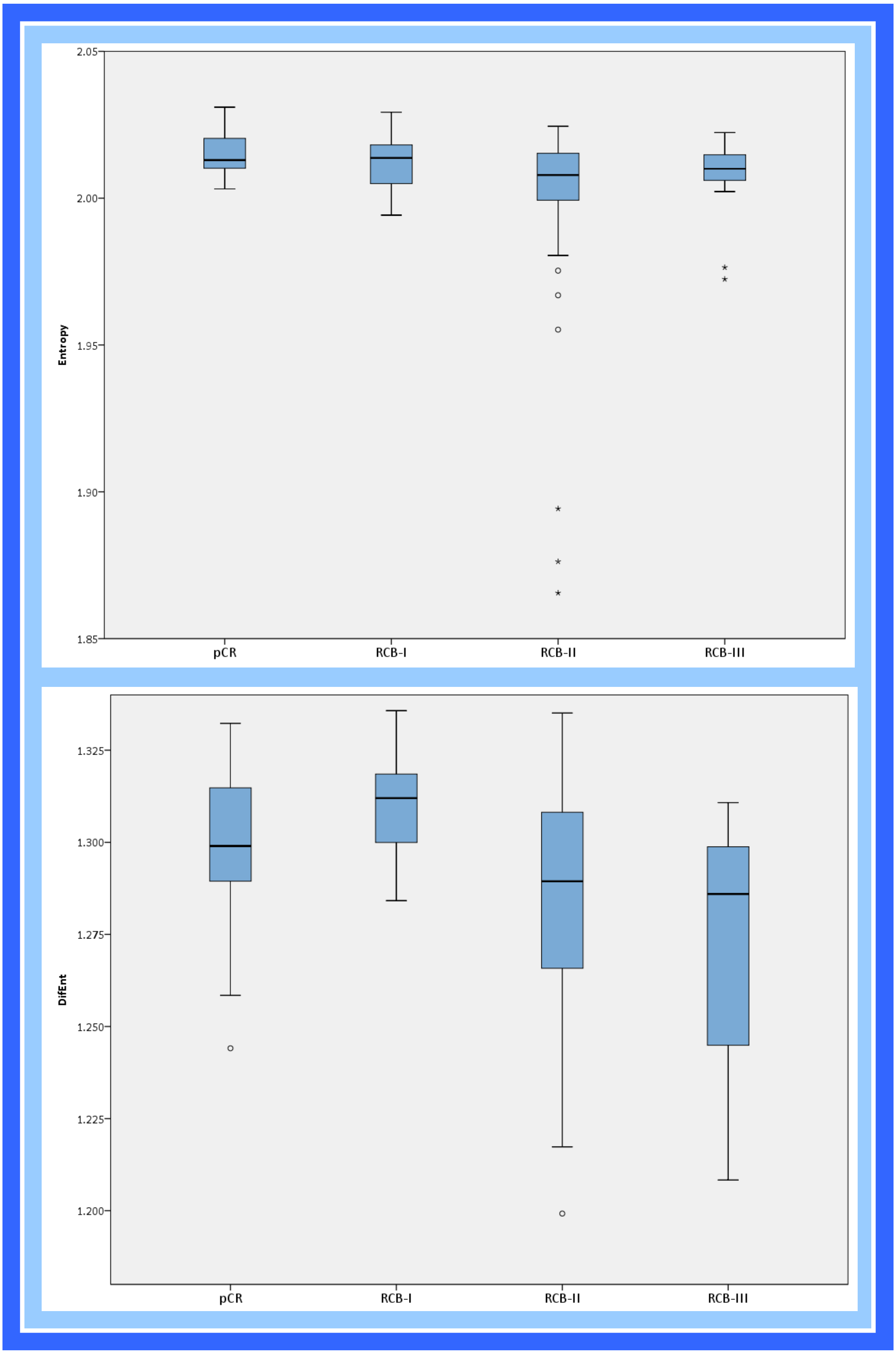


Figure 7.61- Baseline only values for each residual cancer burden index category for lesion regions of interest. Box-whisker plots are displayed for both entropy (upper) and difference entropy (lower) features

From Figure 7.61, it is clear that there is a significant difference in the entropy and difference entropy features between the pathological complete response cases (pCR) and the moderate and extensive residual disease categories (RCB-II and RCB-III, respectively). Considering the absolute values of the features in Figure 7.61, the pCR category lesions have a higher entropy and difference entropy value as measured on the baseline examination, suggesting that these lesions are more heterogenous than the lesions which respond poorly to therapy.

		pCR		*RCB-I*		*RCB-II*		*RCB-III*	
Normal	AngScMom	0.0256	(0.0192)	0.0478	(0.0274)	0.0304	(0.0235)	0.0206	(0.0141)
	DifEnt	1.29	(1.29)	1.30	(1.26)	1.28	(1.29)	1.30	(1.31)
	Entropy	1.84	(1.86)	1.76	(1.81)	1.82	(1.84)	1.91	(1.95)
	SumEnt	1.42	(1.43)	1.36	(1.39)	1.41	(1.42)	1.45	(1.48)
Lesion	AngScMom	0.0107	(0.0107)	0.0108	(0.0108)	0.0111	(0.0109)	0.0109	(0.0108)
	DifEnt	1.29	(1.30)	1.31	(1.31)	1.28	(1.29)	1.27	(1.29)
	Entropy	2.01	(2.01)	2.01	(2.01)	2.00	(2.01)	2.01	(2.01)
	SumEnt	1.50	(1.50)	1.49	(1.49)	1.48	(1.49)	1.49	(1.49)

Table 7.48- Absolute texture feature values in each response category for each region of interest (3sigfigs). Figures provided in brackets are median values.

The difference entropy parameter appears to perform well, also demonstrating a significant difference between RCB-I compared with RCB-II and RCB-III categories on the baseline images (Figure 7.60).

7.4 DISCUSSION

This work has demonstrated initial promise in the use of texture analysis in detecting early response to chemotherapy, albeit within a small cohort of patients.

All texture analysis was performed by matching slices between baseline and interim examinations and comparing texture feature values between the examinations and correlating with final response to treatment, as measured using both traditional RECIST criteria and the state-of-the-art RCB scoring method [163].

Texture feature changes did not correlate well with RECIST criteria, as measured on final imaging examinations. RECIST is a relatively crude measure of tumour response to therapy derived from World Health Organisation guidelines developed in the 1980's. The RECIST criterion was scored by a breast MR expert by assessing the change in dimensions of tumour between baseline and final MR examination. While some significant differences were obtained in the lesion, these were also accompanied by significant changes in the texture features as measured in normal regions of interest and therefore the results were not felt to be a true reflection of changes within the lesion. It is possible that some changes may occur within normal tissue in response to chemotherapeutic treatment, and as no significant differences were measured within background noise between baseline and interim examinations it is unlikely that such changes are merely due to scan-scan differences. As the change in contralateral breast density can be a biomarker for recurrence and cancer prevention in hormone therapy, it is possible that the textural changes occurring within normal tissue are real and could potentially link with outcome measures. However these changes could not be linked within the scope of this study and further investigation would be required to identify such trends. An attempt was also made to link the absolute feature values as measured on the baseline examination with the ultimate response to treatment. There were no links between these data found.

The same analysis process was then followed, but using a difference measure of response- the RCB score. This technique is a pathological technique which assesses a number of factors, including lesion cellularity, dimensions and axillary burden in order to provide a quantitative histopathological assessment. This measure showed statistically significant differences in feature values between the baseline and interim examinations, with significant differences in lesion texture features angular second moment, entropy and sum entropy between the pCR and RCB-II and RCB-III categories and also between the RCB-I and RCB-II category. There was also statistical significance measured in the difference in entropy feature between the RCB-I and RCB-III categories. The difference in entropy feature value between baseline and interim was largest for the pCR category, suggesting there was a large change in this feature relative to the other response categories, which is perhaps intuitive in the cohort who will continue to have a complete response to treatment.

When absolute feature value on the baseline image only was considered, it was demonstrated that there appeared to be significant differences in the initial value of angular second moment, entropy and difference entropy between the pCR v RCB-II and RCB-III categories, as well significant differences in the difference entropy features between the RCB-I v RCB-II and RCB-III categories. Considering the absolute feature values of these categories, the entropy and difference entropy features of the pCR category are actually highest compared with the other groups, suggesting that the lesions that will progress to have a complete response to chemotherapy are more heterogenous than those who will have minimal response, where lesions are relatively more homogeneous. This is in agreement with findings that higher tumour grade (and therefore increased heterogeneity within the tumour) is associated with a better response to neoadjuvant chemotherapy, although a poorer overall disease-free prognosis [164, 165]. This provides an exciting preliminary result that potentially one initial examination can indicate whether the patient will go on to have a good or poor response to chemotherapy and potentially be used in management of patients in terms of therapeutic or surgical planning.

These results demonstrate confounding results when comparing results with RECIST and RCB. RECIST provides a radiological assessment of lesion response to treatment which is limited by technical specifications, reader experience and image quality. Despite still being used routinely in many centres, it was not designed for looking at neoadjuvant therapy for breast lesions but was rather a general technique designed to assess response to drugs. The RCB method is a more quantitative breast and neoadjuvant specific approach based on true pathological tumour characteristics, rather than the appearance of a tumour on the imaging modality that can be confounded by other factors such as inflammatory response, fibrosis [161] as well as factors relating directly to the acquisition process such as patient compliance and technical failure. The RCB method also provides real measures of patient prognosis and therefore is a very useful metric in patient management. Previous studies have demonstrated that while MRI will agree with final pathological size assessment around 70% of the time [166], it can over- or under- estimate the residual tumour in around a third of patients [161, 166]. Such errors in size estimates will therefore have a significant impact on the final response assessment as defined by RECIST [162] which is based entirely on lesion dimension measurements. In a study performed in 2013, it was reported that RECIST criteria agreed with the RCB score of response in less than 20% of cases (Cohens Kappa= 0.38) [167].

This is in agreement with our findings where we found that 4 patients were classified as having minimal residual disease using the RECIST criteria, when the ultimate RCB score was RCB-II, suggesting moderate residual disease. Similarly, one patient with RCB-I score (minimal residual disease on pathology) was classified as having stable disease using RECIST.

This study is the first we are aware of using texture analysis to correlate with ultimate response as assessed using the RCB score system. Ahmed et al. used texture analysis in the prediction of response to chemotherapy for breast cancer patients based on MRI images using a similar technique to this work

[168]. However, response was dichotomised into only two categories- either responders or non-responders using a cut-off threshold of volume changes greater than or less than 50% compared to pre-treatment, respectively. This is again returning to a RECIST- type classification of response and therefore is not designed for, nor does it represent the underlying pathology. Interestingly, this group report increased lesion heterogeneity to be linked with a reduced chemotherapeutic response, whereas our results suggest the opposite. The underlying rationale for this is unclear and warrants further investigation.

Our study considered not only the lesion itself but background and normal regions of interest in order to ensure that any measures and conclusions made on the basis of changes in lesion feature values could be concluded to be real and not due to scan-scan variations. We have therefore reported throughout the comparisons in lesion, normal and background regions to provide a measure of how certain the results are.

The main limitation within this study was the relatively small patient cohort. While there were only 49 patients included in this study, RCB information was available for only 40 of them. This was attributable to a number of patients that had to be excluded for a variety of reasons in order to maintain integrity of the data, however the study provides a foundation for using texture analysis to correlate with RCB scores that warrants further investigation. There was a wide variety of treatment regimes used, which further complicates analysis and could provide another avenue for further investigation with increased patient numbers to assess different treatment regimens as well as ultimate response categories. Due to the small number of patients, no further sub-categorisation of data was possible and these results provide only a preliminary investigation into this technique in patients. It should be noted, however, that HER2 positive and TNBC are the two key indicators for neoadjuvant chemotherapy.

The use of the RCB scoring system is still relatively new, and not widely used in clinical applications, however it has demonstrated a superior correlation with survival compared to other measures [163] and therefore is likely to become more widely utilised in the future. Correlation of these findings with those obtained on imaging is a very novel area and one with many potential applications. It is a more useful measure than traditional comparisons with RECIST criteria and therefore the correlation of texture measures with this parameter is one that provides an exciting and important avenue for further investigation.

CHAPTER SUMMARY: This chapter presents results correlating texture analysis with various measures of response to chemotherapy. There were no correlations found between RECIST response criteria and texture analysis findings, however the RCB score categories appeared to have significantly different entropy and difference entropy features. There is also preliminary findings indicating that absolute baseline texture feature values may be useful in prediction of ultimate treatment response outcome, however this requires further investigation.

CHAPTER EIGHT: CONCLUSIONS AND FUTURE WORK

The aim of this thesis was to investigate how useful the image processing tool, texture analysis, could be in breast MRI examinations. This has been addressed from a number of angles and results presented in previous chapters. This, final, chapter considers the overall outcome from the work and looks to the future and where the findings could lead.

8.1 PROJECT SUMMARY

By constructing a breast-mimicking phantom, incorporating texture phantoms in the form of 4 different grades of reticulated foam, an assessment was made of the ability of texture analysis to differentiate between test objects that were visually undistinguishable on resulting MR images, and also to compare the effect of acquisition parameter changes and different acquisition conditions. Scanning was performed across two different scanner platforms with different field strengths and receiver RF breast coils. We were able to demonstrate that in all cases, texture analysis could reliably differentiate between different grades of foam, despite there being no difference visually obvious in the images acquired, thus showing promise for the technique of texture analysis. The results were consistent across both scanner platforms and acquisition protocols, with the most critical factor in the outcome of texture analysis being in the spatial resolution of the acquired images. This is to be expected as TA is a statistically based technique and therefore is reliant on pixel size. There were no correlations found in the outcome of texture analysis and the measured signal to noise ratio and therefore texture analysis was considered to be attractive for further work considering patient data, without any changes to the existing imaging protocols, and the benefits of thus using currently clinically acquired MRI data.

A retrospective patient data set was then considered by using data from patients gathered over a number of months. Pathology data was also acquired in order to link findings with the classifications of the lesions. Firstly, comparison was made of malignant lesions with those that were found to be benign (despite a malignant appearance on MR dynamic imaging) and normal tissue. Texture analysis demonstrated with a 100% accuracy the ability to differentiate between the three tissue types (malignant, benign, normal) and therefore showed initial promise for clinical utility. Consideration was then made of classification of the malignant lesions into their respective subtypes. Different methods of lesion classification were used- their histological subtype, receptor status as well as the full molecular subtype classification. Good classification accuracies were obtained for all of these, with the exception of using the lesion's HER2 status, which resulted in very poor classification accuracies. Entropy features appeared to demonstrate significant differences in feature value, however gave low classification accuracies- potentially due to the lower number of sample points. As has been evident from publications cited in earlier chapters, entropy has been reported to be of significant interest in texture analysis of cancer imaging as it provides a marker of internal tumoural heterogeneity. Within our research, this provided expected measures of lesion morphological characteristics, such as demonstrating high entropy within triple-negative cancers, and therefore it is hypothesised that there may be an indirect link between underlying growth patterns and the pattern of contrast uptake within lesions, although the mechanism is unclear.

This work was then extended in order to provide an assessment of how the technique would perform clinically by recruiting an independent test set. The results from this study were less convincing (as is sometimes the case with such test sets in general), with molecular subtyping resulting in lower classification accuracies, although histological subtype and grading results showed promise. The number of patients recruited was relatively small, however, and this ultimately was due to the inclusion criteria and limited timeframes for the study over which we had little control. By combining the entire dataset and utilising an external validation method, we demonstrated that increasing the sample numbers in the training set could, potentially, result in an increase in the classification accuracies and therefore these results hold initial preliminary promise for further investigative work to be carried out.

Entropy was demonstrated to be a critical feature in describing lesions, and this was reinforced when the work carried out on the retrospective patient set was repeated using the Max-Min texture analysis method. This Max-Min technique was developed in order to address computational issues, which are largely irrelevant now. However, by considering the number of extremes in the pixel intensity distribution (i.e. a measure of heterogeneity), a similar trend to the previous work on the same patient cohort was found, and the measure of heterogeneity appeared to link in with the results from the entropy feature using the COM model. As subtypes of cancer were largely well differentiated using this technique also, this corroborates the hypothesis that entropy is one of the main classifying features in the categorisation of breast cancer subtypes.

In the drive for patient-tailored treatments, and monitoring of therapy, texture analysis was finally applied to this theme, with a specific focus on early response to neoadjuvant chemotherapy for breast cancer. Texture analysis was performed on both baseline and interim examinations to identify if early changes could be detected that would ultimately predict the outcome of therapy. These results were found not to correlate at all with the traditional RECIST criteria of response, yet, reassuringly, correlated very well with the pathological outcome measure of residual cancer burden (RCB score). As this is a measure of pathological assessment of the actual tumour response, rather than the appearance on a scan, the link between the TA results and this definitive pathology (with its implications for outcome) is very exciting in the drive towards assessment of early response to chemotherapy. TA appeared to be able to categorise between those patients who would have a complete or good response and those who would not. While our cohort was relatively small, there were also preliminary findings that suggested that absolute baseline texture features may also prove to be a link with the final outcome of treatment, although this warrants more exploration both in prospective clinical series and in the setting of randomised controlled trials of neoadjuvant therapy.

8.2 FINAL CONCLUSIONS

The results of the work carried out in this thesis confirm that texture analysis, as initially reported within the literature, is an exciting image processing tool with extensive potential. Within the area of cancer, and specifically breast cancer, the advantages of an image-based technique that can differentiate between molecular subtypes and particularly can provide an early indication as to eventual response to chemotherapy given prior to surgery could have powerful implications for diagnosis, individualisation of treatment and the management of women with breast cancer more generally. However, as has been demonstrated by this work, there is still a significant effort required in order to develop the techniques into a robust clinical tool.

Preliminary indications suggest that there could be potential for classifications between different types of breast cancer as well as an indicator of early response to treatment, however in order to make any definitive conclusions, it is essential to significantly increase patient numbers in order to make reliable, robust conclusions.

The research within this thesis has demonstrated that texture analysis is a useful research tool within the area of breast MRI. While it is not ready for clinical use at present until further development is carried out, it has provided great preliminary promise which certainly warrants further investment in this area.

8.3 FUTURE WORK

As was perhaps to be expected, this thesis has reported on findings that have ultimately led to a desire to extend this work beyond the scope of the initial research proposed.

As has been alluded to earlier, it is essential that the numbers included within the patient cohort are significantly increased. While initial results prove promising, in order to create a robust classification model, further data is required to develop this. It would also be useful to assess the effect of lesion size and region of interest size in order to ascertain the limit(s) in the technique. It is hypothesised that this will be limited by pixel size, rather than absolute lesion size, however this, as yet, remains untested.

Also extending the work of this thesis further, an increase in the NAC population is imperative in order to further investigate the link between the residual cancer burden scores and the textural changes between baseline and interim examination. The work carried out within the scope of this research indicates preliminary findings that texture could provide a measure of response that appears to link with the pathological outcome of the treatment and therefore this could potentially prove to be an exciting and essential area of research in improving patient management.

There has been much work reported within the literature that the influence of peri-tumoural stroma has influence on the long term prognosis of patient outcome. Therefore, extending this work to consider not only the tumour itself, but also tumour periphery and into the surrounding stroma would prove a useful further investigation to carry out. Due to the excellent depiction of soft tissue on MR, this may be the most suitable modality to perform such research.

Finally, it has been well established that patients with BRCA1 and BRCA2 gene mutations are at a higher risk of developing breast cancer, particularly at a young age. Using a large scale study, most likely across multiple institutions, it would be fascinating to consider the fibroglandular tissue of patients within this group compared to women with standard population risk of breast cancer development. Using texture analysis, could it be possible to identify fundamental textural differences between the fibroglandular tissues of these two groups? While such research would require, most likely, UK wide collaboration to obtain sufficient patient numbers, a study investigating this issue would truly be exciting.

Texture analysis of MR images in primary breast cancer is clearly at an early stage at what may prove to be a very exciting and clinically relevant tool to improve the management of women with breast cancer.

REFERENCES

1. Fan F. and Thomas PA., *Common Benign Conditions of Low Concern*, in *Breast Cancer and its Precursor Lesions*, Thomas PA., Editor. 2011, Humana Press: New York.
2. Seeley RR, Stephens TD, and Tate P, *Reproductive System*, in *Anatomy & Physiology*. 2003, McGraw-Hill: New York. p. 1015-1060.
3. Morris EA., *The Normal Breast*, in *Breast MRI*, Morris EA. and Liberman L., Editors. 2005, Springer: New York. p. 23-44.
4. Lyrl. *MenstrualCycle*. 2007 [cited 17/12/13]; Available from: http://commons.wikimedia.org/wiki/File:MenstrualCycle2.png.
5. Vogel PM., Georgiade NG., Fetter BF., Vogel FS., and McCarty KS., *The correlation of histological changes in the human breast with the menstrual cycle.* Americal Journal of Pathology, 1981. **104**(1): p. 23-34.
6. Ramakrishnan R., Khan SA., and Badve S., *Morphological changes in breast tissue with menstrual cycle.* Modern Pathology, 2002. **15**(12): p. 1348-1356.
7. Potten CS., Watson RJ., Williams GT., Tickle S., Roberts SA., Harris M., and Howell A., *The effect of age and menstrual cycle upon proliferative activity of the normal human breast.* British Journal of Cancer, 1988. **58**: p. 163-170.
8. Milligan D., Drife JO., and Short RV., *Changes in breast volume during normal menstrual cycle and after orla contraceptives.* British Medical Journal, 1975. **4**: p. 494-496.
9. Fowler PA., Casey CE., Cameron GG., Foster MA., and Knight CH., *Cyclic changes in composition and volume of the breast during the menstrual cycle, measured by magnetic resonance imaging.* BJOG: An International Journal of Obstetrics and Gynaecology, 1990. **97**(7): p. 595-602.
10. Thomas PA., *Benign Conditions Associated with a Risk for the Subsequent Development of Cancer*, in *Breast Cancer and its Precursor Lesions*, Thomas PA., Editor. 2011, Humana Press: New York.
11. Cancer Research UK. *CancerStats. Breast Cancer-UK*. CancerStats [Online Publication] 2009 May 2009 [cited 2010 15/2/10]; Available from: http://publications.cancerresearchuk.org/WebRoot/crukstoredb/CRUK_PDFs/CSBREA03.pdf.
12. Michell M., Reznek RH., and Husband JE., eds. *Breast Cancer*. Contemporary Issues in Cancer Imaging. 2010, Cambridge University Press: Cambridge.
13. Khamis ZI., Shab ZJ., and Sang Q-A., *Active roles of tumour stroma in breast cancer metastasis.* International Journal of Breast Cancer, 2012. **Article ID:574025**: p. 10pages.
14. Page DL, Steel CM, and Dixon JM, *Carcinoma in situ and patients at high risk of breast cancer.*, in *ABC of Breast Diseases*, Dixon JM, Editor. 2001, BMJ Publishing: London.
15. Hesketh R., *Introduction to Cancer Biology*. 2013, Cambridge: Cambridge University Press.

16. Kneeshaw PJ, Lowry M, Manton D, Hubbard A, Drew PJ, and Turnbull LW, *Differentiation of malignant from benign disease associated with screening detected microcalcifications using dynamic contrast enhanced magnetic resonance imaging* The Breast, 2006. **15**(1): p. 29-38.
17. Schipper H and Baum M, *Pathophysiology*, in *Fast facts- Breast Cancer*. 2005, Health Press Ltd: Oxford. p. 27-48.
18. Tavassoli FA and Devilee P, eds. *World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Breast and Female Genital Organs*. 2003, IARC Press: Lyon.
19. Eheman CR., Shaw KM., Ryerson AB., Miller JW., Ajani UA., and White MC., *The changing incidence of in situ and invasive ductal and lobular breast carcinomas: United State, 1999-2004.* Cancer Epidemiology, Biomarkers and Prevention, 2009. **18**: p. 1763-1769.
20. Qureshi H.S., Linden M.D., Divine G., and Raju U.B., *E-cadherin status in breast cancer correlates with histologic type but does not correlate with established prognostic parameters.* Anatomic Pathology, 2006. **125**: p. 377-385.
21. Goldhirsch A, Ingle JN, Gelber RD, Coates AS, Thurlimann B, Senn H-J, and Members, *Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2009.* Annals of Oncology, 2009. **20**: p. 1319-1329.
22. Haupt B, Ro JY, and Schwartz MR, *Basal-like breast carcinoma. A phenotypically distinct entity.* Archives of Pathology and Laboratory Medicine, 2010. **134**: p. 130-133.
23. Keller PJ, Lin AF, Arendt LM, Klebba I, Jones AD, Rudnik JS, DiMeo TA, Gilmore H, Jefferson DM, Graham RA, Naber SP, Schnitt S, and Kuperwasser C, *Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines.* Breast Cancer Research, 2010. **12**: p. R87.
24. Altman DG and Bland JM, *Diagnostic Tests 1: Sensitivity and Specificity.* British Medical Journal 1994. **308**: p. 1552.
25. Altman DG and Bland JM, *Diagnostic Tests 2: Predictive Values.* British Medical Journal 1994. **309**: p. 102.
26. Kolb TM, Lichy J, and Newhouse JH, *Comparison of the performance of screening mammography, physical examination, and breast ultrasound and evaluation of factors that influence them: An analysis of 27,825 patient evaluations.* Radiology, 2002. **225**: p. 165-175.
27. Berg WA, Gutierrez L, NessAiver MS, Carter WB, Bhargavan M, Lewis RS, and Ioffe OB, *Diagnostic accuracy of mammography, clinical examination, US and MR imaging in preoperative assessment of breast cancer.* Radiology, 2004. **233**: p. 830-849.
28. Donegan WL, *Evaluation of a palpable breast mass.* New England Journal of Medicine, 1992. **327**(13): p. 937-942.
29. Baines CJ, Miller AB, and Bassett AA, *Physical Examination: its role as a single screening modality in the Canadian national breast screening study.* Cancer, 1989. **63**: p. 1816-1822.

30. Barton MB, Harris R, and Fletcher SW, *Does this patient have breast cancer? The screening clinical breast examination : should it be done? how?* The Journal of the American Medical Association, 1999. **282**(13): p. 1270-1280.
31. Pisano ED, Gatsonis C, Hendrick RE, Yaffe MJ, Baum JK, Acharrya S, Conant EF, Fajardo LL, Bassett L, D'Orsi CJ, Jong RA, and Rebner M, *Diagnostic performance of digital versus film mammography for breast cancer screening.* New England Journal of Medicine, 2005. **353**(17): p. 1773-1783.
32. Warner E, Plewes DB, Hill KA, Causer PA, Zubovits JT, Jong RA, Cutrara MR, DeBoer G, Yaffe MJ, Messner SJ, Meschino WS, Piron CA, and Narod SA, *Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography and clinical examination.* The Journal of the American Medical Association, 2004. **292**(11): p. 1317-1325.
33. Brem RF, *The future of breast cancer diagnosis: molecular breast imaging.* Mayo Clinic Proceedings, 2005. **80**(1): p. 17-18.
34. Liberman L, Abramson AF, Squires FB, Glassman JR, Morris EA, and Dershaw DD, *The breast imaging and reporting data system: positicve predictive value of mammographic features and final assessment categories.* American Journal of Roentgenology, 1998. **171**(1): p. 35-40.
35. Price JL and Gibbs NM, *The relationship between microcalcification and in situ carcinoma of the breast.* Clinical Radiology, 1978. **29**(4): p. 447-452.
36. Homer MJ, Safaii H, Smith TJ, and Marchant DJ, *The relationship of mammographic microcalcification to histological malignancy: radiologic-pathologic correlation.* American Journal of Roentgenology, 1989. **153**(6): p. 1187-1189.
37. Carney PA, Miglioretti DL, Yankaskas BC, Kerlkowske K, Rosenberg R, Rutter CM, Geller BM, Abraham LA, Taplin SH, Dignan M, Cutter G, and Ballard-Barbash R, *Individual and combined effects of age, breast density and hormone replacement therapy use on the accuracy of screening mammography.* Annals of Internal Medicine, 2003. **138**: p. 168-175.
38. Berrington de Gonzalez A and Reeves G, *Mammographic screening before age 50 years in the UK: comparison of the radiation risks with the mortality benefits.* British Journal of Cancer, 2005. **93**: p. 590-596.
39. Wallis MG., Moa E., Zanca F., Leifland K., and Danielsson M., *Two-view and single-view tomosynthesis versus full-field digital mammography: high-resolution x-ray imaging observer study.* Radiology, 2012. **262**: p. 788-796.
40. Teh W and Wilson ARM, *The role of ultrasound in breast cancer screening. A consensus statement by the European Group for breast cancer screening.* European Journal of Cancer, 1998. **34**(4): p. 449-450.
41. Anderson ME, Soo MS, Bentley RC, and Trahey GE, *The detection of microcalcifications with medical ultrasound.* Journal of the Acoustical Society of America, 1997. **101**(1): p. 29-39.

42. Kuhl C, *The current status of breast MR imaging. Part I- choice of technique, image interpretation, diagnostic accuracy and transfer to clinical practice.* Radiology, 2007. **244**(2): p. 356-378.
43. Vassiou K, Kanavou T, Vlychou M, Poultsidi A, Athanasiou E, Arvanitis DL, and Fezoulidis IV, *Characterization of breast lesions with CE-MR multimodal morphological and kinetic analysis: comparison with conventional mammography and high-resolution ultrasound.* European Journal of Radiology, 2009. **70**: p. 69-76.
44. Lehman C, Gatsonis C, Kuhl CK, Hendrick RE, Pisano ED, Hanna L, Peacock S, Smazal SF, Maki DD, Julian TB, DePeri ER, Bluemke DA, and Schnall MD, *MRI of the contralateral breast in women with recently diagnosed breast cancer.* New England Journal of Medicine, 2007. **353**(13): p. 1295-1303.
45. Turnbull L, Brown S, Harvey I, Olivier C, Drew P, Napp V, Hanby A, and Brown J, *Comparative effectiveness of MRI in breast cancer (COMICE) trial: a randomised controlled trial.* The Lancet, 2010. **375**: p. 563-571.
46. Moy L, Elias K, Patel V, Lee J, Babb JS, Toth HK, and Mercado CL, *Is breast MRI helpful in the evaluation of inconclusive mammographic findings.* American Journal of Roentgenology, 2009. **193**: p. 986-993.
47. Kuhl CK., Klaschik S., Mielcareck P., Gieseke J., Wardelmann E., and Schild H., *Do T2-weighted pulse sequences help with the differential diagnosis of enhancing lesions in dynamic breast MRI?* Journal of Magnetic Resonance Imaging, 1999. **9**: p. 197-196.
48. Kuhl C, Mielcareck P, Klaschik S, Leutner C, Wardelmann E, Gieseke J, and Schild H, *Dynamic breast MR imaging: are signal intensity time course data useful for differential diagnosis of enhancing lesions?* Radiology, 1999. **211**: p. 101-110.
49. Kaiser WA and Zeitler E, *MR Imaging of the breast: faast imaging sequences with and without Gd-DTPA. Preliminary observations.* Radiology, 1989. **170**: p. 681-686.
50. Orel SG and Schnall MD, *MR imaging of the breast for the detection, diagnosis and staging of breast cancer.* Radiology, 2001. **220**: p. 13-30.
51. Kuhl CK, Schrading S, Bieling HB, Wardelmann E, Leutner CC, Koenig R, Kuhn W, and Schlid HH, *MRI for diagnosis of pure ductal carcinoma in situ: a prospective observational study.* The Lancet, 2007. **370**(9586): p. 485-492.
52. American College of Radiology, *Breast imaging reporting and data system atlas (BI-RADS atlas)*. 2003, Reston, VA: American College of Radiology.
53. Erguvan-Dogan E, Whitman GJ, Kushwaha AC, Phelps MJ, and Dempsey PJ, *BI-RADS-MRI: A Primer.* American Journal of Roentgenology, 2006. **187**(2): p. W152-W160.
54. Solin LJ, Orel SG, Hwang WT, Harris EA, and Schnall MD, *Relationship of breast magnetic resonance imaging to outcome after breast-conservation treatment with radiation for women with early-stage invasive breast carcinoma or ductal carcinoma in situ.* Journal of Clinical Oncology, 2008. **26**(3): p. 386-391.

55. Christov K, Chew KL, Ljung B-M, Waldman FM, Duarte LA, Goodson WH, Smith HS, and Mayall BH, *Proliferation of normal breast epithelial cells as shown by in vivo labelling with bromodeoxyuridine.* American Journal of Pathology, 1991. **138**(6): p. 1371-1377.

56. Olsson H, Jernstrom H, Alm P, Kreipe H, Ingvar C, Jonsson P-E, and Ryden S, *Proliferation of the breast epithelium in relation to menstrual cycle phase, hormonal use and reproductive factors.* Breast Cancer Research and Treatment, 1996. **40**(2): p. 187-196.

57. Kuhl CK, Bieling HB, Gieseke J, Kreft BP, Sommer T, Lutterbey G, and Schild HH, *Healthy premenopausal breast parenchyma in dynamic contrast-enhanced MR imaging of the breast: normal contrast medium enhancement and cyclical-phase dependency.* Radiology, 1997. **203**: p. 137-144.

58. Hendrick RE, *Contrast Agents in Breast Magentic Resonance Imaging*, in *Breast MRI*. 2008, Springer: New York. p. 113-134.

59. Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, Larsson HBW, Lee T-Y, Mayr NA, Parker GJM, Port RE, Taylor J, and Weisskoff MR, *Estimating kinetic parameters from dynamic contrast-enhances T1 weighted MRI of a diffusable tracer: standardized quantities and symbols.* Journal of Magnetic Resonance Imaging, 1999. **10**: p. 223-232.

60. Schabel MC, Morrell GR, Oh KY, Walczak CA, Barlow RB, and Neumayer LA, *Pharmokinetic mapping for lesion characterisation in dynamic breast MRI.* Journal of Magnetic Resonance Imaging, 2010. **31**: p. 1371-1378.

61. Tudorica LA., Oh KY., Roy N., Kettler MD., Chen Y., Hemmingson SL., Afzal A., Grinstead JW., Laub G., Li X., and Huang W., *A feasible high spatiotemporal resolution breast DCE-MRI protocol for clinical settings.* Magnetic Resonance Imaging, 2012. **30**(9): p. 1257-1267.

62. Mori S. and Barker P.B., *Diffusion magnetic resonance imaging: its principle and applications.* The Anatomical Record, 1999. **257**(3): p. 102-109.

63. Charles-Edwards E.M. and deSouza N.M., *Diffusion-weighted magnetic resonance imaging and its application to cancer.* Cancer Imaging, 2006. **6**: p. 135-143.

64. Sinha S, Lucas-Quesada FA, Sinha U, DeBruhl N, and Bassett LW, *In vivo diffusion-weighted MRI of the breast: potential for lesion characterization.* Journal of Magnetic Resonance Imaging, 2002. **15**: p. 693-704.

65. Woodhams R, Matsunaga K, Iwabuchi K, Kan S, Hata H, Kuranami M, Watanabe M, and Hayakawa K, *Diffusion weighted imaging of malignant breast tumors. The usefulness of apparent diffusion coefficient (ADC) value and ADC map for the detection of malignant breast tumours and evaluation of cancer extension.* Journal of Computer Assisted Tomography, 2005. **29**(5): p. 644-649.

66. Marini C, Iacconi C, Gianelli M, Cilotti A, Moretti M, and Bartolozzi C, *Quantitative diffusion-weighted imaging in the differential diagnosis of breast lesion.* European Radiology, 2007. **17**: p. 2646-2655.

67. Peters NH, V.K., van den Bosch AM, Luijten PR, Mali WP, Bartels LW, *Quantitative diffusion weighted imaging for differentiation of benign and malignant breast lesions: the influence of the choice of b-values.* Journal of Magnetic Resonance Imaging, 2010. **31**: p. 1100-1105.
68. Guo Y, Cai Y-Q, Cai Z-L, Gao Y-G, An N-Y, Ma L, Mahankali S, and Gao J-H, *Differentiation of clinically benign and malignant breast lesions using diffusion-weighted imaging.* Journal of Magnetic Resonance Imaging, 2002. **16**: p. 172-178.
69. Sinha S and Sinha U, *Recent advances in breast MRI and MRS.* NMR in Biomedicine, 2009. **22**: p. 3-16.
70. Baek H-M, Yu HJ, Chen J-H, Nalcioglu O, and Su M-Y, *Quantitative correlation between 1H MRS and dynamic contrast-enhanced MRI of human breast cancer.* Magnetic Resonance Imaging, 2008. **26**: p. 523-531.
71. Bonini RH, Zeotti D, Saraiva LA, Trad CS, Filho JM, Carrara HH, de Andrade JM, Santos AC, and Muglia VF, *Magnetization transfer ratio as a predictor of malignancy in breast lesions: preliminary results.* Magnetic Resonance in Medicine, 2008. **59**: p. 1030-1034.
72. Calas MJG., Almeida RMVR., Gutfielen B., and Pereira WCA., *Interobserver concordance int he BI-RADS classification of breast ultrasound exams.* Clinics, 2012. **67**(2): p. 185-189.
73. Skaane P., Engedal K., and Skjennald A., *Interobserver variation in the interpretation of breast imaging comparison of mammography, ultrasonography, and both combined in the interpretation of palpable noncalcified breast masses.* Acta Radiologica, 1997. **38**(4): p. 497-502.
74. Redondo A., Comas M., Macia F., Ferrer F., Murta-Nascimento C., Maristany MT., Molins E., Sala M., and Castells X., *Inter- and intra-radiologist variability in the BI-RADS assessment and breast density categroies for screening mammograms.* British Journal of Radiology, 2012. **85**: p. 1465-1470.
75. Uematsu T., Kasami M., Yuen S., Igarashi T., and Nasu H., *Comparison of 3- and 1.5-T dynamic breast MRI for visualisation of spiculated masses previously identified using mamography.* American Journal of Roentgenology, 2012. **198**: p. W611-W617.
76. Ko JM., Nicholas MJ., Mendel JB., and Slanetz P., *Prospective assessment of computer-aided detection in interpretation of screening mammography.* American Journal of Roentgenology, 2006. **187**: p. 1483-1491.
77. Taplin SH., Rutter CM., and Lehman C., *Testing the effect of computer assisted detection on interpretive performance in screening mammography.* American Journal of Roentgenology, 2006. **187**: p. 1475-1482.
78. Gur D., Sumkin JH., Rockette HE., Ganott M., Hakim C., Hardesty L., Poller WR., Shah R., and Wallace L., *Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system.* Journal of the National Cancer Institute, 2004. **96**(3): p. 185-190.

79. Balleyguier C., Kinkel K., Fermanian J., Malan S., Djen G., Taourel P., and Helenon O., *Computer-aided detection (CAD) in mammography: does it help the junior or senior radiologist?* European Journal of Radiology, 2005. **54**.

80. Shen W-C., Chang R-F., Moon WK., Chou Y-H., and Huang C-S., *Breast ultrasound computer-aided diagnosis using BI-RADS features.* Academic Radiology, 2007. **14**: p. 928-939.

81. Meeuwis C., Van de Ven SM., Stapper G., Gallardo AMF., van den Bosch MAAJ., Mali WPTM., and Veldhuis WB., *Computer-aided detection (CAD) for breast MRI: evaluation of efficacy at 3.0T.* European Radiology, 2010. **20**: p. 522-528.

82. Pan J., Dogan BE., Carkaci S., Santiago L., Arribas E., Cantor SB., Wei W., Stafford RJ., and Witman GJ., *Comparing performance of the CADstream and the DynaCAD breast MRI CAD systems.* Journal of Digital Imaging, 2013. **26**: p. 971-976.

83. Arazi-Kleinman T., Causer PA., Jong RA., Hill K., and Warner E., *Can breast MRI computer aided detection (CAD) improve radiologist accuracy for lesions detected at MRI screening and recommended for biopsy in a high risk population.* Clinical Radiology, 2009. **64**: p. 1166-1174.

84. Renz DM., Bottcher J., Diekmann F., Poellinger A., Maurer MH., Pfeil A., Streitparth F., Colletini F., Bick U., Hamm B., and Fallenberg EM., *Detection and classification of contrast-enhancing masses by a fully automatic computer-assissted diagnosis system for breast MRI.* JMRI, 2012. **35**: p. 1077-1088.

85. Nie K, Chen J-H, Yu HJ, Chu Y, Nalcioglu O, and S. M-Y, *Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI.* Academic Radiology, 2008. **15**(12): p. 1513-1525.

86. Materka A, *What is texture?*, in *Texture Analysis for Magnetic Resonance Imaging*, D.M. Hajek M, Materka A, Lerski R, Editor. 2006, Med4publishing: Prague. p. 11-43.

87. Lerski RA, Straughan K, Schad LR, Boyce D, Bluml S, and Zuna I, *MR image texture analysis- an approach to tissue characterisation.* Magnetic Resonance Imaging, 1993. **11**: p. 873-887.

88. Haralick RM, Shanmugam K, and Dinstein I, *Textural features for image classification.* IEEE Transactions on Systems, Man and Cybernetics, 1973. **3**(6): p. 610-621.

89. Haralick RM, *Statistical and structural approaches to texture.* Proceedings of the IEEE, 1979. **67**(5): p. 786-804.

90. Tourassi GD, *Journey toward computer-aided diagnosis: role of image texture analysis.* Radiology, 1999. **213**: p. 317-320.

91. Kjaer L, Ring P, Thomsen C, and Henriksen O, *Texture analysis in quantitative MR imaging. Tissue characterisation of normal brain and intracranial tumours at 1.5T.* Acta Radiologica, 1995. **36**(2): p. 127-135.

92. Nedelec J-F, Yu O, Chambron J, and Macher J-P, *Texture analysis of the brain: from animal models to human applications.* Dialogues in Clinical Neuroscience, 2004. **6**(2): p. 227-233.

93. Yu O, Mauss Y, Namer IJ, and Chambron J, *Existence of contralateral abnormalities revealed by texture analysis in unilateral intractable hippocampal epilepsy.* Magnetic Resonance Imaging, 2001. **19**: p. 1305-1310.
94. Lerski R, *Clinical applications of texture analysis*, in *Texture Analysis for Magnetic Resonance Imaging*, H. M, et al., Editors. 2006, Med4publishing: Prague. p. 151-191.
95. Zacharaki EI, Wang S, Chawla S, Soo Yoo D, Wolf R, Melhem ER, and Davatzikos C, *Classification of brain tumour type and grade using MRI texture and shape in a machine learning scheme.* Magnetic Resonance in Medicine, 2009. **62**: p. 1609-1618.
96. Jirak D, Dezortova M, Taimr P, and Hajek M, *Texture analysis of human liver.* Journal of Magnetic Resonance Imaging, 2002. **15**: p. 68-74.
97. Link TM, Majumdar S, Augat P, Lin JC, Newitt D, Lu Y, Lane NE, and Genany HK, *In vivo high resolution MRI of the calcaneus: differences in trabecular structure in osteoporosis patients.* Journal of Bone and Mineral Research, 1998. **13**(7): p. 1175-1182.
98. Langenberger H, Shimza Y, Windischberger C, Grampp S, Berg A, Ferlitsch K, and Moser E, *Bone homogeneity factor: an advanced tool for the assessment of osteoporotic bone structure in high resolution magnetic resonance images.* Investigative Radiology, 2003. **38**(7): p. 467 - 472.
99. Petrosian A, C.H.-P., Helvie MA, Goodsitt MM, Adler DD, *Computer-aided diagnosis in mammography: classification of mass and normal tissue by texture analysis.* Physics in Medicine and Biology, 1994. **29**: p. 2273-2288.
100. Sivaramakrishna R, Powell KA, Lieber ML, Chilcote WA, and Shekhar R, *Texture analysis of lesions in breast ultrasound images.* Computerized Medical Imaging and Graphics 2002. **26**: p. 303-307.
101. Gupta R and Undrill PE, *The use of texture analysis to delineate suspicious masses in mammography.* Physics in Medicine and Biology, 1995. **40**: p. 835-855.
102. Garra BS, Krasner BH, Horii SC, Ascher S, Mun SK, and Zeman RK, *Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis.* Ultrasonic Imaging, 1993. **15**: p. 267-285.
103. Sinha S, Lucas-Quesada FA, DeBruhl ND, Sayre J, Farria D, Gorczyca DP, and Bassett LW, *Multifeature analysis of Gd-enhanced MR images of breast lesions.* Journal of Magnetic Resonance Imaging, 1997. **7**(6): p. 1016-1026.
104. Gibbs P and Turnbull LW, *Textural analysis of contrast-enhanced MR images of the breast.* Magnetic Resonance in Medicine, 2003. **50**: p. 92-98.
105. Chen W, Giger ML, Li H, Bick U, and Newstead GM, *Volumetric texture analysis of breast lesions on contrast enhanced magnetic resonance images.* Magnetic Resonance in Medicine, 2007. **58**: p. 562-571.

106. Gatenby RA, Grove O, and Gillies RJ, *Quantitative imaging in cancer evolution and oncology.* Radiology, 2013. **269**(1): p. 8-15.
107. Win T, Miles KA, Jones AM, Ganeshan B, Shastry M, Endozo R, Meagher M, Shortman RI, Wan S, Kayani I, Ell PJ, and Groves AM, *Tumour heterogeneity and permeability as measured on the CT component of PET/CT predict survival in patients with non-small cell lung cancer.* Clinical Cancer Research, 2013. **19**: p. 3591-3599.
108. Davnall F, Yip CS, Ljungqvist G, Selmi M, Hg F, Sanghera B, Ganeshan B, Miles KA, Cook GJ, and Goh V, *Assessment of tumour heterogeneity: an emerging imaging tool for clinical practice?* Insights Imaging, 2012. **3**: p. 573-589.
109. Ganeshan B, Strukowska O, Skogen K, Young R, Chatwin C, and Miles K, *Heterogeneity of focal breast lesions and surrounding tissue assessed by mammographic texture analysis: prelimiary evidence of an associated with tumour invasion and estrogen receptor status.* Frontiers in Oncology, 2011. **1**: p. 33.
110. Holli K, Laaperi AL, Harrison L, Luukkaala T, Toivonen T, Ryymin P, Dastidar P, Soimakallio S, and Eskola H, *Characterisation of breast cancer types by texture analysis of magnetic reonance images.* Academic Radiology, 2010. **17**(2): p. 135-141.
111. Lerski RA, Schad LR, Luypaert R, Amorison A, Muller RN, Mascaro l, Ring P, Spisni A, Zhu X, and Bruno A, *Multicentre magnetic resonance texture analysis trial using reticulated foam test objects.* Magnetic Resonance Imaging, 1999. **17**(7): p. 1025-1031.
112. Magnet Lab. *MRI: A Guided Tour*. 2013 [cited 30/12/2013]; Available from: http://www.magnet.fsu.edu/education/tutorials/magnetacademy/mri/index.html.
113. McRobbie DW, Moore EA, Graves MJ, and Prince MR, *Seeing is believing: introduction to image contrast*, in *MRI from Picture to Proton*. 2007, University Press: Cambridge. p. 27-45.
114. Oxford English Dictionary. *"texture, n."* online version: June 2011 [cited 2011, 7th July]; Available from: http://www.oed.com/view/Entry/200031.
115. Davies ER, *Introduction to Texture Analysis*, in *Handbook of Texture Analysis*, Mirmehdi M, Xie X, and Suri J, Editors. 2008, Imperial College Press: London. p. 1-32.
116. Tamura H, Mori S, and TYamawaki T, *Textural features corresponding to visual perception.* IEEE Transactions on Systems, Man and Cybernetics, 1978. **8**(6): p. 460-473.
117. Tuceryan M, *Texture Analysis*, in *The Handbook of Pattern Recognition and Computer Vision*, Chen CH, Pau LF, and Wang PSP, Editors. 1999, World Scientific Publishing Company: Singapore. p. 207-249.
118. Nailon WH, *Texture analysis methods for medical image characterisation*, in *Biomedical Imaging*, Mao Y, Editor. 2010, InTech.

119. Landy MS and G. N, *Visual Perception of Texture*, in *The Visual Neurosciences, Vol 1*, Chalupa LM and W. JS, Editors. 2004, Massachusetts Institute of Technology: USA. p. 1106-1118.
120. Bhalerao A and Reyes-Aldasoro CC, *Volumetric texture description and discrimnant feature selection for MRI.* Lecture Notes in Computer Science, 2003. **2809**: p. 573-584.
121. Castellano G, Bonilha L, Li LM, and Cendes F, *Texture analysis of medical images.* Clinical Radiology, 2004. **59**: p. 1061-1069.
122. Mitchell OR, Myers CR, and B. W, *A Max-Min Measure for Image Texture Analysis.* IEEE Transactions on Computers, 1977. **26**(4): p. 408-414.
123. Julesz B, *Experiments in the visual perception of texture.* Scientific American, 1975. **232(4)**: p. 34-43.
124. Materka A, *MaZda User Manual*. 2006, http://www.eletel.p.lodz.pl/programy/mazda/download/mazda_manual.pdf: Institute of Electronics, Technical University of Lodz.
125. Technical University of Lodz, *MaZda*. 2010: Lodz, Poland.
126. Szczpinski PM, Strzelecki M, Materka A, and Klepaczko A, *MaZda- a software package for image texture analysis.* Computer Methods and Programs in Biomedicine, 2009. **94**(1): p. 66-76.
127. Strzlecki M, Materka A, and Szczypinski P, *MaZda*, in *Texture Analysis for Magnetic Resonance Imaging*, D.M. Hajek M, Materka A, Lerski R, Editor. 2006, Med4Publishing: Prague. p. 107-113.
128. Technical University of Lodz, *B11*. 2010: Lodz.
129. Materka A, *B11*. 2004, University of Lodz, Institute of Electronics. p. MRI texture data analysis and classification.
130. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, and Witten IH, *The WEKA Data Mining Software: An Update.* SIGKDD Explorations, 2009. **11**(1): p. 10-18.
131. IBM Corp., *IBM SPSS Statistics for Windows, Version 21.0*. Released 2012, IBM Corp.: Armonk, NY.
132. Mayerhoefer ME, Szomolanyi P, Jirak D, Berg A, Materka A, Dirisamer A, and Trattnig S, *Effects of magnetic resonance image interpolation on results of texture based pattern classification.* Investigative Radiology, 2009. **44**(7): p. 405-411.
133. Jirak D, Dezortova M, and Hajek M, *Phantoms for texture analysis of MR images. Long term and multi-center study.* Medical Physics, 2004. **31**(3): p. 616-622.
134. Herlidou-Meme S, Constans JM, Carsin B, Olivie D, Eliat PA, Nadal-Desbarats L, Gondry C, Le-Rumeur E, Idy-Peretti I, and de-Certaines JD, *MRI texture analysis on texture test objects, normal brain and intracranial tumours.* Magnetic Resonance Imaging, 2003. **21**: p. 989-993.
135. Mayerhoefer ME, Szomolanyi P, Jirak D, Materka A, and Trattnig S, *Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination.* Medical Physics, 2009. **36**(4): p. 1236-1243.

136. Schad LR and Lundervold A, *Influence of resolution and signal to noise ratio on image texture*, in *Texture Analysis for Magnetic Resonance Imaging*, D.M. Hajek M, Materka A, Lerski R, Editor. 2006, Med4Publishing: Prague. p. 129-149.
137. Collewet G, Strzelecki M, and Mariette F, *Influenece of MRI acquisition protocols and image intensity normalisation methods on texture classification.* Magnetic Resonance Imaging, 2004. **22**: p. 81-91.
138. Materka A, Strzelecki M, Lerski R, and Schad L, *Toward automatic feature selection of texture test objects for magnetic resonance imaging*, in *RECPAD 2000, 11th Portugese Conference on Pattern Recognition*. 2000, RECPAD: Porto.
139. Lerski RA and Schad LR, *The use of reticulated foam in texture test objects for magnetic resonance imaging.* Magnetic Resonance Imaging, 1998. **16**(9): p. 1139-1144.
140. Rakow-Penner R, Daniel B, Yu H, Sawyer-Glover A, and Glover GH, *Relaxation times of breast tissue at 1.5T and 3.0T measured using IDEAL.* Journal of Magnetic Resonance Imaging, 2006. **23**: p. 87-91.
141. Mann RM, Kuhl CK, Kinkel K, and Boetes C, *Breast MRI: guidelines from the European Society of Breast Imaging.* European Radiology, 2008. **18**: p. 1307-1318.
142. Radjenovic A, Dall BJ, Ridgway JP, and Smith MA, *Measurement of pharmacokinetic parameters in histologically graded invasive breast tumours using dynamic contrast enhanced MRI.* British Journal of Radiology 2008. **81**: p. 120-128.
143. Su M-Y, Yu HJ, Carpenter PM, McLaren CE, and Nalcioglu O, *Pharmacokinetic parameters analysed from MR contrast enhancement kinetics of multiple malignant and benign breast lesions detected in the same patients.* Technology in Cancer Research and Treatment, 2005. **4**: p. 255-264.
144. Hoffmann U, Brix G, Knopp MV, HeB T, and Lorenz WJ, *Pharmacokinetic mapping of the breast: a new method for dynamic MR mammography.* Magnetic Resonance in Medicine, 1995. **33**: p. 506-514.
145. Szczpinski PM, *Mazda 3D Editor*. 2007, Technical University of Lodz, Institute of Electronics. p. MRI Analysis Software.
146. Szczpinski PM, *Personal Communication: Software developers teaching visit*. 2010: Lodz, Poland.
147. Duda RO, Hart PE, and Stork DG, *Pattern Classification*. 2nd ed. 2001, USA: John Wiley & Sons Inc.
148. Greenberg R., Skornick Y., and K. O., *Management of breast fibroadenomas.* Journal of General Internal Medicine, 1998. **13**(9): p. 640-645.
149. Bland JM and Altman DG, *Statistical methods for assessing agreement between two methods of clinical measurement.* Lancet, 1986. **i**: p. 307-310.
150. Eliat PA, Olivie D, Saikali S, Carsin B, Saint-Jalmes H, and de Certaines JD, *Can dynamic contrast-enhanced magnetic resonance imaging combined with texture analysis differentiate malignant glioneuronal tumours from*

other glioblastoma? Neurology Research International, 2012. **2012**: p. Article ID 195176.

151. Berg WA, Blume JD, Cormack JB, Mendelson EB, Lehrer D, Bohm-Velez M, Pisano ED, Jong RA, Evans WP, Morton MJ, Mahoney MC, Larsen LH, Barr RG, Farria DM, Marques HS, and B. K, *Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer.* Journal of the American Medical Association, 2008. **299**(18): p. 2151-2163.
152. Stoutjesdijk MJ, Boetes C, Jager GJ, Beex L, Bult P, Hendriks JHCL, Laheij RJF, Massuger L, Die LE, Wobbes T, and B. JO, *Magnetic resonance imaging and mammography in women with a hereditary risk of breast cancer.* Journal of the National Cancer Institute, 2001. **93**(14): p. 1095-1102.
153. Skogen K, Ganeshan B, Good T, Critchley G, and Miles KA, *Imaging hetereogeneity in gliomas using texture analysis.* Cancer Imaging, 2011. **11**(Special No A): p. S113.
154. Bernasconi A, Antel SB, Collins DL, Bernasconi N, Olivier A, Dubeau F, Pike GB, Andermann F, and A. DL, *Texture Analysis and Morphological Processing of Magnetic Resonance Imaging Assist Detection of Focal Cortical Dysplasia in Extra-Temporal Partial Epilepsy.* Annals of Neurology, 2001. **49**(6): p. 770-775.
155. Lerski R, Smith MJ, Morley P, Barnett E, Mills PR, Watkinson G, and MacSween RNM, *Discrimnant Analysis of Ultrasonic Texture Data in Diffuse Alcoholic Liver Disease: Fatty liver and cirrhosis.* Ultrasonic Imaging, 1981. **3**(164-172).
156. Lerski R, Barnett E, Morley P, Mills PR, Watkinson G, and MacSween RNM, *Computer Analysis of Ultrasonic Signals in Diffuse Liver Disease.* Ultrasound in Medicine & Biology, 1979. **5**: p. 341-350.
157. Giat P, Berger G, Laugier P, and Abouelkaram S, *Application of a Max-Min Algorithm on Ultrasonic Signals for Tissue Texture Analysis*, in *14th Annual Conference of the IEEE*. 1992: Paris, France. p. 2104-2105.
158. Faul F., Erdfelder E., Lang A-G., and Buckhner A., *G*Power 3: A felxible statistical power analysis program for the social, behavioural and biomedical sciences.* Behaviour Research Methods, 2007. **39**: p. 175-191.
159. Mackay A., Weigelt B., Grigoriadis A., Kreike B., Natrajan R., A'Hern R., Tan D., Dowsett M., Ashworth A., and Reis-Filho J., *Microarray-based class discovery for molecular classification of breast cancer: analysis of interobserver agreement.* Journal of the National Cancer Institute, 2011. **103**: p. 662-673.
160. Schott AF. and H. DF., *Defining the benefits of neoadjuvant chemotherapy for breast cancer.* Journal of Clinical Oncology, 2012. **30**(15): p. 1747-1749.
161. Yeh R, Slanetz P, Kopans DB, Rafferty E, Georgian-Smith D, Moy L, Halpern E, Moore R, Kuter I, and Taghian A, *Prospective comparison of mammography, sonography, and MRI in patients undergoing neoadjuvant chemotherapy for palpable breast cancer.* American Journal of Roentgenology, 2005. **184**(3): p. 868-877.

162. Eisenhauer E.A., Therasse P., Bogaerts J., Schwartz L.H., Sargent D., Ford R., Dancey J., Arbuck S., Gwyther S., Mooney M., Rubinstein L., Shankar L., Dodd L., Kaplan R., Lacombe D., and Verweij J., *New response evaluation criteria in solid tumours: Revised RECIST guidelines (version 1.1).* European Journal of Cancer, 2009. **45**: p. 228-247.
163. Symmans W.F., Peintinger F., Hatzis C., Rajan R., Kuerer H., Valero V., Assad L., Poniecka A., Hennessy B., Green M., Buzdar A.U., Singletary S.E., Hortobagyi G.N., and Pusztai L., *Measurement of residual cancer burden to predict survival after neoadjuvant chemotherapy.* Journal of Clinical Oncology, 2007. **25**(28): p. 4414-4422.
164. Fisher ER., Wang J., Bryant J., Fisher B., Mamounas E., and Wolmark N., *Pathobiology of preoperative chemotherapy: findings from the national surgical adjuvant breast and bowel (NSABP) protocol B-18.* Cancer, 2002. **95**(4): p. 681-695.
165. Liedtke C., Hatzis C., Symmans W.F., Desmedt C., Haibe-Kains B., Valero V., Kuerer H., Hortobagyi G.N., Piccart-Gebhart M., Sotiriou C., and Pusztai L., *Genomic grade index is associated with response to chemotherapy in patients with breast cancer.* Journal of Clinical Oncology, 2009. **27**(19): p. 3185-3191.
166. Kim HJ., Im YH>, Han BK., Choi N., Lee J., Kim JH., Choi YL., Ahn JS., Nam SJ., Park YS., Choe YH., Ko YH., and Yang JH., *Accuracy of MRI for estimating residual tumour size after neoadjuvant chemotherapy in locally advanced breast cancer: Relation to response patterns on MRI.* Acta Oncologica, 2007. **46**: p. 996-1003.
167. Romero A., Garcia-Saenz JA., Fuentes-Ferrer M., Garcia-Asenjo JAL., Furio V., Roman JM., Moreno A., de la Hoya M., Diaz-Rubio E., Martin M., and C. T., *Correlation between response to neoadjuvant chemotherapy and survivial in locally advanced breast cancer patients.* Annals of Oncology, 2013. **24**: p. 655-661.
168. Ahmed A., Gibbs P., Pickles M., and Turnbull L., *Texture analysis in assessment and prediction of chemotherapy response in breast cancer.* Journal of Magnetic Resonance Imaging, 2013. **38**: p. 89-101.

APPENDIX A: MATHEMATICAL DESCRIPTION OF TEXTURE FEATURES

STATISTICAL BASED METHODS

Histogram-Based Features

For an image histogram, with intensity levels defined from *i=1…N_g*, a normalised histogram vector, *p(i)*, is defined and the following histogram-based features derived.

$Mean, \mu$	$\sum_{i=1}^{N_g} ip(i)$
$Variance, \sigma^2$	$\sum_{i=1}^{N_g} (i-\mu)^2 p(i)$
$Skewness, \mu_3$	$\sigma^{-3}\sum_{i=1}^{N_g} (i-\mu)^3 p(i)$
$Kurtosis, \mu_4$	$\sigma^{-4}\sum_{i=1}^{N_g} (i-\mu)^4 p(i)-3$

Co-Occurrence Matrix Features

The co-occurrence matrix is defined as the estimated joint probability $p_{d\theta}(i,j)$ of two pixels a distance *d* (d=1,2,3,4,5) apart in given direction θ (θ=0°,45°,90°,135°) having co-occurring values of *i* and *j*, in an image (*f(x,y))* with intensity levels defined from *i=1…N_g*. The co-occurrence matrix is defined with the *(i,j)* entry given by the number of times that

$$f(x_1, y_1) = i \ , \ f(x_2, y_2) = j$$
$$(x_2, y_2) = (x_1, y_1) + (d \cos\theta \sin\theta)$$

Co-occurrence matrix parameters are then derived as below.

Angular Second Moment	$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)^2$	
Contrast	$\sum_{n=0}^{N_g-1} n^2 \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)$	$\lvert i-j \rvert = n$
Correlation	$\dfrac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} ijp(i,j) - \mu_x\mu_y}{\rho_x\rho_y}$	
Sum of Squares	$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} (i-\mu_x)^2\, p(i,j)$	
Inverse Difference Moment	$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{1}{1+(i-j)^2}\, p(i,j)$	
Entropy	$-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\log(p(i,j))$	
Sum Average	$\sum_{i=1}^{2N_g} ip_{x+y}(i)$	
Sum Entropy	$-\sum_{i=1}^{2N_g} p_{x+y}(i)\log(p_{x+y}(i))$	$p_{x+y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)$
Sum Variance	$\sum_{i=1}^{2N_g} (i - SumAverage)^2\, p_{x+y}(i)$	$i+j=k; \quad k=2,3,\ldots,2N_g$
Difference Variance	$\sum_{i=0}^{N_g-1} (i-\mu_{x-y})^2\, p_{x-y}(i)$	$p_{x-y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)$
Difference Entropy	$-\sum_{i=1}^{N_g} p_{x-y}(i)\log(p_{x-y}(i))$	$\lvert i-j \rvert = k; \quad k=0,1,\ldots,N_g-1$

Absolute Gradient Features

MaZda considers the 3×3 pixel neighbourhood of an image pixel *x(i,j),* such that

A	B	C	D	E
F	G	H	I	J
K	L	*x(i,j)*	N	O
P	Q	R	S	T
U	V	W	Y	Z

$$ABSV(i,j) = \sqrt{(R-H)^2 + (N-L)^2}$$

For the matrix of M elements containing the gradient values for pixels in the region of interest (ROI), gradient features are defined as below.

Mean Absolute Gradient (GrMean)	$\frac{1}{M}\sum_{i,j\in ROI} ABSV(i,j)$
Gradient Variance (GrVariance)	$\frac{1}{M}\sum_{i,j\in ROI}(ABSV(i,j)-GrMean)^2$
Gradient Skewness	$\frac{1}{\left(\sqrt{GrVariance}\right)^3}\frac{1}{M}\sum_{i,j\in ROI}(ABSV(i,j)-GrMean)^3$
Gradient Kurtosis	$\frac{1}{\left(\sqrt{GrVariance}\right)^4}\frac{1}{M}\sum_{i,j\in ROI}(ABSV(i,j)-GrMean)^4-3$

Run-Length Matrix Features

For an image with intensity levels defined from *i=1…*N_g, the run-length matrix *p(i,j)* is defined as the number of times there is a run of length *j* having grey-level *i*, with N_r being the number of runs.

Short run emphasis inverse moment	$\left(\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{p(i,j)}{j^2}\right)\Big/C$
Long run emphasis moments	$\left(\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} j^2 p(i,j)\right)\Big/C$
Grey level non-uniformity	$\left(\sum_{i=1}^{N_g}\left(\sum_{j=1}^{N_r} p(i,j)\right)^2\right)\Big/C$
Run length non-uniformity	$\left(\sum_{j=1}^{N_r}\left(\sum_{i=1}^{N_g} p(i,j)\right)^2\right)\Big/C$
Fraction of image in runs	$\left(\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j)\right)\Big/\left(\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} jp(i,j)\right)$
Coefficient, C	$\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j)$

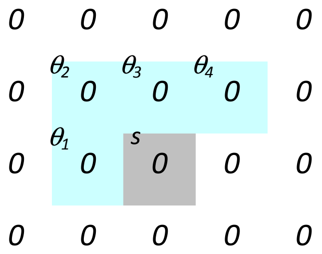
MODEL BASED METHODS

Auto-Regressive Model Features

The autoregressive model assumes that pixel intensity at a given site (*s*) is a weighted sum (f_s) of neighbouring pixel intensities and independent distributed noise (e_s) as described by

$$f_s = \sum_{r \in N_s} \theta_r f_r + e_s$$

The model parameters are described by vectors (θ) and MaZda implements an autoregressive model utilising four parameter vectors (θ_1, θ_2, θ_3, θ_4) and the standard deviation of the noise, σ, to describe the texture in the immediately surrounding area of the pixel *s*.



The parameters are then estimated by minimising the sum of the squared error to give the following linear equations, which are solved for each ROI of interest.

$$\hat{\theta} = \left(\sum_s w_s w_s^T \right)^{-1} \left(\sum_s w_s f_s \right)$$

$$w_s = col\left[f_i, i \in N_s \right]$$

$$\sigma^2 = \frac{1}{\sqrt{N}} \sum_s \left(f_s - \hat{\theta} w_s \right)^2$$

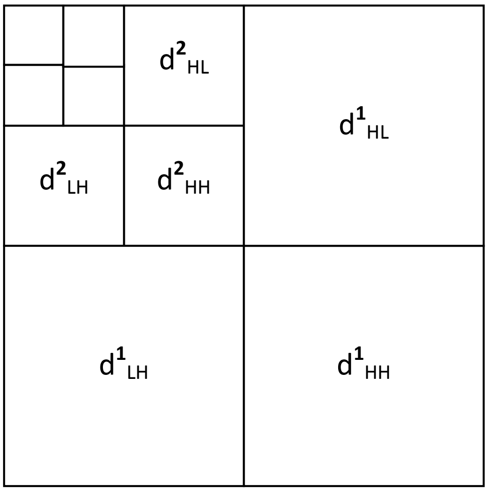
TRANSFORM BASED METHODS

Wavelet Transform Features

The wavelet transform is a method of separating data into high (H) and low (L) frequency components and considering these components with resolutions matched to their scale. MaZda utilises a Haar wavelet transform and considers only the energy feature at each sub-band (there are four image sub-bands at each scale- d_{HH}, d_{HL}, d_{LH}, d_{LL}).

Energy can be calculated for any ROI with a number of pixels given by *n*, at any scale by:

$$E_{subband,scale} = \sum_{x,y\in ROI}\left(d_{x,y}^{subband}\right)^{2} \Big/ n$$



APPENDIX B: PROGRAM CODING USED FOR MAX-MIN ANALYSIS

ROI GENERATION

This code was written using Matlab 2012(a) (MathWorks; Massachusetts, USA). The program loads in a selected DICOM image and generates a 10×10 square ROI in an area selected by the user. A 10×10 text file of pixel values is generated and then resized to be 100×1 for calculation of Max-Min texture values.

```
% clear all memory to initialise program
clc
clear
close all
imtool close all
cd('C:\Documents and Settings\swaugh\Desktop\MATLAB
info\Images');

% read in image, display and display copy of image for ROI to be
drawn on
importfile=uigetfile('*.*');
image= dicomread(importfile);
imshow(image,[]);
image2=imresize(image,2);
imshow(image2,[]);
title('Original with ROI');
set(gcf, 'Position', get(0, 'ScreenSize'));
set(gcf,'name','testing, testing','numbertitle','off');

% create ROI and display pixel values
roi=impoint(gca);
wait(roi);
position=ans;
xpos=position(1);
ypos=position(2);
roi=imrect(gca,[xpos ypos 10 10]);
mask=roi.createMask();
pixelvalues=image2(mask);
matrix=reshape(pixelvalues,10,10);
display(matrix);

% writes pixel value matrix to txt file
m=double(matrix);
n= reshape(m,100,1)
fid=fopen('matrix.txt','wt');
fprintf(fid, [repmat('%g\t', 1, size(n,2)-1) '%g\n'], n.');
fclose(fid)
```

CALCULATION OF MAX-MIN TEXTURE FEATURES

This code was written using Intel Fortran (Intel Corp., California, USA) and X-Code 3.2.5 (Apple Inc.; California, USA) for Mac OSX. The program prompts the user for a threshold value and an increment value in order to calculate number of maximum and minimum extreme values at a number of threshold values. Ratios are then calculated in order to determine the textural description features.

```
        PROGRAM maxmin
cDEC$ FIXEDFORMLINESIZE:132
c       A program to attempt the implementation of the
c       Max-Min texture method of Mitchell et al (IEEE trans Comp
1977)
c       on clinical image files.
c
C       The Texture Analysis is performed firstly on horizontal
lines with
C       10x10 ROIs. The matrix is then flipped through 90 degrees
so that the
C       the procedure can be repeated thus doing vertical lines.
c
        Character*64 fname
        Character*3 rdata(100),rd
        real irdata(100),sdata(100),ird,T,Tinc,Nrat,krdata(100)
        integer ncount(5),IJ(100)
c
        DATA IJ(1),IJ(2),IJ(3),IJ(4),IJ(5)/1,11,21,31,41/
        DATA IJ(6),IJ(7),IJ(8),IJ(9),IJ(10)/51,61,71,81,91/
        DATA IJ(11),IJ(12),IJ(13),IJ(14),IJ(15)/2,12,22,32,42/
        DATA IJ(16),IJ(17),IJ(18),IJ(19),IJ(20)/52,62,72,82,92/
        DATA IJ(21),IJ(22),IJ(23),IJ(24),IJ(25)/3,13,23,33,43/
        DATA IJ(26),IJ(27),IJ(28),IJ(29),IJ(30)/53,63,73,83,93/
        DATA IJ(31),IJ(32),IJ(33),IJ(34),IJ(35)/4,14,24,34,44/
        DATA IJ(36),IJ(37),IJ(38),IJ(39),IJ(40)/54,64,74,84,94/
        DATA IJ(41),IJ(42),IJ(43),IJ(44),IJ(45)/5,15,25,35,45/
        DATA IJ(46),IJ(47),IJ(48),IJ(49),IJ(50)/55,65,75,85,95/
        DATA IJ(51),IJ(52),IJ(53),IJ(54),IJ(55)/6,16,26,36,46/
        DATA IJ(56),IJ(57),IJ(58),IJ(59),IJ(60)/56,66,76,86,96/
        DATA IJ(61),IJ(62),IJ(63),IJ(64),IJ(65)/7,17,27,37,47/
        DATA IJ(66),IJ(67),IJ(68),IJ(69),IJ(70)/57,67,77,87,97/
        DATA IJ(71),IJ(72),IJ(73),IJ(74),IJ(75)/8,18,28,38,48/
        DATA IJ(76),IJ(77),IJ(78),IJ(79),IJ(80)/58,68,78,88,98/
        DATA IJ(81),IJ(82),IJ(83),IJ(84),IJ(85)/9,19,29,39,49/
        DATA IJ(86),IJ(87),IJ(88),IJ(89),IJ(90)/59,69,79,89,99/
        DATA IJ(91),IJ(92),IJ(93),IJ(94),IJ(95)/10,20,30,40,50/
        DATA IJ(96),IJ(97),IJ(98),IJ(99),IJ(100)/60,70,80,90,100/
c
```

```
        write(*,'(A\)') 'enter filename:  '
        read(*,'(A\)') fname
c
        open(3,file=fname)
        Do 1 I=1,100
1       read(3,'(A3)') rdata(I)
        pause 'read data'
        print *, rdata
c
c       convert to real log data
        DO 10 I=1,100
        rd=rdata(I)
        read (rd, '(g6.0)') irdata(I)
        irdata(I)=alog(irdata(I)+1.0)
10      continue
        print *, irdata
c
c       Count initial peaks (before smoothing)
400     ncount(1)=0
        DO 30 ILINE=0,9
        DO 20 I=1,8
        IF
(irdata(ILINE+I)<irdata(ILINE+I+1).AND.irdata(ILINE+I+1)>irdata(I
LINE+I+2)) THEN
        ncount(1)=ncount(1)+1
        endif
20      continue
30      continue
c
        print *, ncount
c
        write(*,'(A\)') 'enter starting T value:  '
        read(*,*) Tstart
        write(*,'(A\)') 'enter T increment value:  '
        read(*,*) Tinc
c
c       Smooth some data
        sdata(1)=irdata(1)
        T=Tstart
        DO 31 ILINE=0,9
        DO 21 I=1,8
        IF (sdata(ILINE+I)<(irdata(ILINE+I+1)-T/2)) THEN
        sdata(ILINE+I+1)=(irdata(ILINE+I+1)-T/2)
        ELSE IF ((irdata(ILINE+I+1)-
T/2)<sdata(ILINE+I).AND.sdata(ILINE+I)<irdata(ILINE+I+1)+T/2)
THEN
        sdata(ILINE+I+1)=sdata(ILINE+I)
        ELSE IF ((irdata(ILINE+I+1)+T/2)<sdata(ILINE+I)) THEN
        sdata(ILINE+I+1)=irdata(ILINE+I+1)+T/2
        endif
21      continue
31      continue
c
c       Count peaks (after first smoothing)
        ncount(2)=0
        DO 32 ILINE=0,9
        DO 22 I=1,8
        IF
(sdata(ILINE+I)<sdata(ILINE+I+1).AND.sdata(ILINE+I+1)>sdata(ILINE
+I+2)) THEN
        ncount(2)=ncount(2)+1
```

```
        endif
22      continue
32      continue
c
c       Perform second smooth and count
        T=Tstart+Tinc
        sdata(1)=irdata(1)
        DO 33 ILINE=0,9
        DO 23 I=1,8
        IF (sdata(ILINE+I)<(irdata(ILINE+I+1)-T/2)) THEN
        sdata(ILINE+I+1)=(irdata(ILINE+I+1)-T/2)
        ELSE IF ((irdata(ILINE+I+1)-
T/2)<sdata(ILINE+I).AND.sdata(ILINE+I)<irdata(ILINE+I+1)+T/2)
THEN
        sdata(ILINE+I+1)=sdata(ILINE+I)
        ELSE IF ((irdata(ILINE+I+1)+T/2)<sdata(ILINE+I)) THEN
        sdata(ILINE+I+1)=irdata(ILINE+I+1)+T/2
        endif
23      continue
33      continue
c
c       Count peaks (after second smoothing)
        ncount(3)=0
        DO 34 ILINE=0,9
        DO 24 I=1,8
        IF
(sdata(ILINE+I)<sdata(ILINE+I+1).AND.sdata(ILINE+I+1)>sdata(ILINE
+I+2)) THEN
        ncount(3)=ncount(3)+1
        endif
24      continue
34      continue
c
c       Perform third smooth and count
        T=Tstart+2*Tinc
        sdata(1)=irdata(1)
        DO 35 ILINE=0,9
        DO 25 I=1,8
        IF (sdata(ILINE+I)<(irdata(ILINE+I+1)-T/2)) THEN
        sdata(ILINE+I+1)=(irdata(ILINE+I+1)-T/2)
        ELSE IF ((irdata(ILINE+I+1)-
T/2)<sdata(ILINE+I).AND.sdata(ILINE+I)<irdata(ILINE+I+1)+T/2)
THEN
        sdata(ILINE+I+1)=sdata(ILINE+I)
        ELSE IF ((irdata(ILINE+I+1)+T/2)<sdata(ILINE+I)) THEN
        sdata(ILINE+I+1)=irdata(ILINE+I+1)+T/2
        endif
25      continue
35      continue
c
c       Count peaks (after third smoothing)
        ncount(4)=0
        DO 36 ILINE=0,9
        DO 26 I=1,8
        IF
(sdata(ILINE+I)<sdata(ILINE+I+1).AND.sdata(ILINE+I+1)>sdata(ILINE
+I+2)) THEN
        ncount(4)=ncount(4)+1
        endif
26      continue
36      continue
```

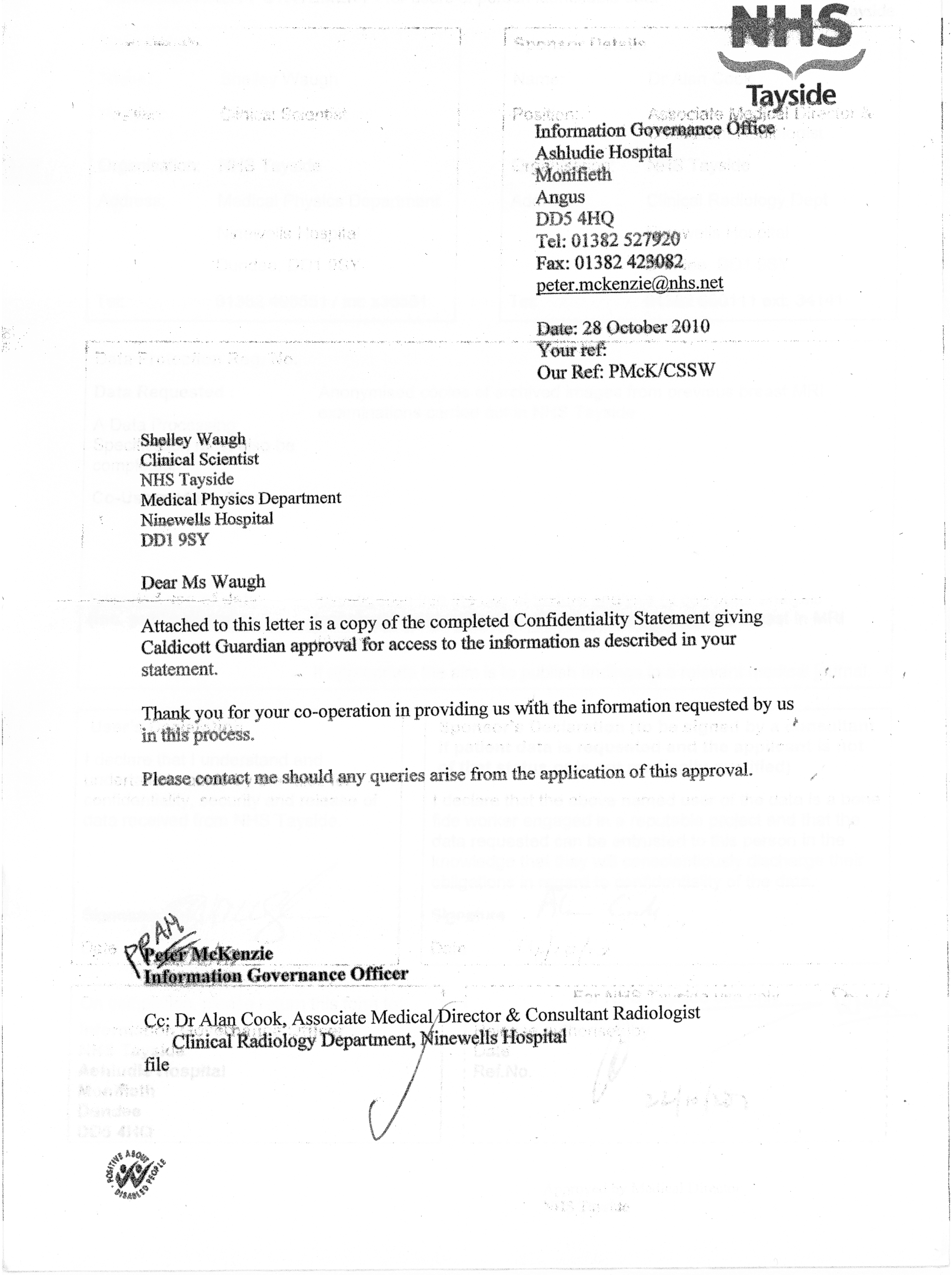
```
c
c       Perform fourth smooth and count
        T=Tstart+3*Tinc
        sdata(1)=irdata(1)
        DO 37 ILINE=0,9
        DO 27 I=1,8
        IF (sdata(ILINE+I)<(irdata(ILINE+I+1)-T/2)) THEN
        sdata(ILINE+I+1)=(irdata(ILINE+I+1)-T/2)
        ELSE IF ((irdata(ILINE+I+1)-
T/2)<sdata(ILINE+I).AND.sdata(ILINE+I)<irdata(ILINE+I+1)+T/2)
THEN
        sdata(ILINE+I+1)=sdata(ILINE+I)
        ELSE IF ((irdata(ILINE+I+1)+T/2)<sdata(ILINE+I)) THEN
        sdata(ILINE+I+1)=irdata(ILINE+I+1)+T/2
        endif
27      continue
37      continue
c
c       Count peaks (after fourth smoothing)
        ncount(5)=0
        DO 38 ILINE=0,9
        DO 28 I=1,8
        IF
(sdata(ILINE+I)<sdata(ILINE+I+1).AND.sdata(ILINE+I+1)>sdata(ILINE
+I+2)) THEN
        ncount(5)=ncount(5)+1
        endif
28      continue
38      continue
C
C       Flip ROI through 90 degrees to count vertical lines
        do 100 I=1,100
        J=ij(I)
100     krdata(I)=irdata(J)
c       repeat all smoothing and counting in vertical direction
c
c       Count initial peaks (before smoothing) adding to
horizontal peaks
        DO 130 ILINE=0,9
        DO 120 I=1,8
        IF
(krdata(ILINE+I)<krdata(ILINE+I+1).AND.krdata(ILINE+I+1)>krdata(I
LINE+I+2)) THEN
        ncount(1)=ncount(1)+1
        endif
120     continue
130     continue
c
        print *, ncount
c
c       Smooth some data
        sdata(1)=krdata(1)
        T=Tstart
        DO 131 ILINE=0,9
        DO 121 I=1,8
        IF (sdata(ILINE+I)<(krdata(ILINE+I+1)-T/2)) THEN
        sdata(ILINE+I+1)=(krdata(ILINE+I+1)-T/2)
        ELSE IF ((krdata(ILINE+I+1)-
T/2)<sdata(ILINE+I).AND.sdata(ILINE+I)<krdata(ILINE+I+1)+T/2)
THEN
        sdata(ILINE+I+1)=sdata(ILINE+I)
```

```
        ELSE IF ((krdata(ILINE+I+1)+T/2)<sdata(ILINE+I)) THEN
        sdata(ILINE+I+1)=krdata(ILINE+I+1)+T/2
        endif
121     continue
131     continue
c
c       Count peaks (after first smoothing)
        DO 132 ILINE=0,9
        DO 122 I=1,8
        IF
(sdata(ILINE+I)<sdata(ILINE+I+1).AND.sdata(ILINE+I+1)>sdata(ILINE
+I+2)) THEN
        ncount(2)=ncount(2)+1
        endif
122     continue
132     continue
c
c       Perform second smooth and count
        T=Tstart+Tinc
        sdata(1)=krdata(1)
        DO 133 ILINE=0,9
        DO 123 I=1,8
        IF (sdata(ILINE+I)<(krdata(ILINE+I+1)-T/2)) THEN
        sdata(ILINE+I+1)=(krdata(ILINE+I+1)-T/2)
        ELSE IF ((krdata(ILINE+I+1)-
T/2)<sdata(ILINE+I).AND.sdata(ILINE+I)<krdata(ILINE+I+1)+T/2)
THEN
        sdata(ILINE+I+1)=sdata(ILINE+I)
        ELSE IF ((krdata(ILINE+I+1)+T/2)<sdata(ILINE+I)) THEN
        sdata(ILINE+I+1)=krdata(ILINE+I+1)+T/2
        endif
123      continue
133      continue
c
c       Count peaks (after second smoothing)
        DO 134 ILINE=0,9
        DO 124 I=1,8
        IF
(sdata(ILINE+I)<sdata(ILINE+I+1).AND.sdata(ILINE+I+1)>sdata(ILINE
+I+2)) THEN
        ncount(3)=ncount(3)+1
        endif
124      continue
134      continue
c
c       Perform third smooth and count
        T=Tstart+2*Tinc
        sdata(1)=krdata(1)
        DO 135 ILINE=0,9
        DO 125 I=1,8
        IF (sdata(ILINE+I)<(krdata(ILINE+I+1)-T/2)) THEN
        sdata(ILINE+I+1)=(krdata(ILINE+I+1)-T/2)
        ELSE IF ((krdata(ILINE+I+1)-
T/2)<sdata(ILINE+I).AND.sdata(ILINE+I)<krdata(ILINE+I+1)+T/2)
THEN
        sdata(ILINE+I+1)=sdata(ILINE+I)
        ELSE IF ((krdata(ILINE+I+1)+T/2)<sdata(ILINE+I)) THEN
        sdata(ILINE+I+1)=krdata(ILINE+I+1)+T/2
        endif
125     continue
135     continue
```

```
c
c       Count peaks (after third smoothing)
        DO 136 ILINE=0,9
        DO 126 I=1,8
        IF
(sdata(ILINE+I)<sdata(ILINE+I+1).AND.sdata(ILINE+I+1)>sdata(ILINE
+I+2)) THEN
        ncount(4)=ncount(4)+1
        endif
126     continue
136     continue
c
c       Perform fourth smooth and count
        T=Tstart+3*Tinc
        sdata(1)=krdata(1)
        DO 137 ILINE=0,9
        DO 127 I=1,8
        IF (sdata(ILINE+I)<(krdata(ILINE+I+1)-T/2)) THEN
        sdata(ILINE+I+1)=(krdata(ILINE+I+1)-T/2)
        ELSE IF ((krdata(ILINE+I+1)-
T/2)<sdata(ILINE+I).AND.sdata(ILINE+I)<krdata(ILINE+I+1)+T/2)
THEN
        sdata(ILINE+I+1)=sdata(ILINE+I)
        ELSE IF ((krdata(ILINE+I+1)+T/2)<sdata(ILINE+I)) THEN
        sdata(ILINE+I+1)=krdata(ILINE+I+1)+T/2
        endif
127     continue
137     continue
c
c       Count peaks (after fourth smoothing)
        DO 138 ILINE=0,9
        DO 128 I=1,8
        IF
(sdata(ILINE+I)<sdata(ILINE+I+1).AND.sdata(ILINE+I+1)>sdata(ILINE
+I+2)) THEN
        ncount(5)=ncount(5)+1
        endif
128     continue
138     continue

        print *, ncount
c
c       calculate count ratios
        DO 39 I=2,5
        Nrat=FLOAT(ncount(I))/FLOAT(ncount(1))
        write(*,*) Nrat
39      continue
c
        write(*,'(A\)') 'again (1) or stop (0)?  '
        read(*,*) iagain
        IF (iagain.EQ.1) THEN
        DO 200 I=1,5
200     Ncount(I)=0
        GOTO 400
        endif
        end
```

APPENDIX C: CALDICOTT APPROVAL



NHS
Tayside
Information Governance Office
Ashludie Hospital
Monifieth
Angus
DD5 4HQ
Tel: 01382 527920
Fax: 01382 423082
peter.mckenzie@nhs.net

Date: 28 October 2010
Your ref:
Our Ref: PMcK/CSSW

Shelley Waugh
Clinical Scientist
NHS Tayside
Medical Physics Department
Ninewells Hospital
DD1 9SY

Dear Ms Waugh

Attached to this letter is a copy of the completed Confidentiality Statement giving Caldicott Guardian approval for access to the information as described in your statement.

Thank you for your co-operation in providing us with the information requested by us in this process.

Please contact me should any queries arise from the application of this approval.

Peter McKenzie
Information Governance Officer

Cc: Dr Alan Cook, Associate Medical Director & Consultant Radiologist
Clinical Radiology Department, Ninewells Hospital
file

NHS
Tayside

CONFIDENTIALITY STATEMENT - for users of person identifiable data

User Details

Name: Shelley Waugh
Position: Clinical Scientist
Organisation: NHS Tayside
Address: Medical Physics Department
Ninewells Hospital
Dundee, DD1 9SY
Tel: 01382 496551 / Int: x36551

Sponsor Details

Name: Dr Alan Cook
Position: Associate Medical Director & Consultant Radiologist
Organisation: NHS Tayside
Address: Clinical Radiology Dept
Ninewells Hospital
Dundee, DD1 9SY
Tel: 01382 660111 ext: 34141

Data Protection Reg. No:

Data Requested : Anonymised copies of archived images from previous breast MRI examinations carried out in NHS Tayside.

A Data Processing Specification must also be completed.

Co-Users of the Data :

Intended use of data (inc. publications) : Investigation into the use of texture analysis (a computer analysis technique) for identification of malignant regions in the breast in MRI images.

If appropriate the aim is to publish findings in a relevant medical journal.

User's Declaration

I declare that I understand and undertake to abide by the rules for confidentiality, security and release of data received from NHS Tayside.

Signature SWaugh
Date 12/10/10

Sponsor's Declaration (to be signed by a consultant if patient data is requested and the applicant is not of that status or is not medically qualified)

I declare that the above named user of the data is a bona fide worker engaged in a reputable project and that the data requested can be entrusted to this person in the knowledge that they will conscientiously discharge their obligations in regard to confidentiality of the data.

Signature Alan Cook
Date 14/10/10

On completion, please return this form to:
Information Governance Officer
NHS Tayside
Ashludie Hospital
Monifieth
Dundee
DD5 4HQ

For NHS Tayside use only
Release authorised by
Date 22/10/2010
Ref.No.

Approved by Medical Director
NHS Tayside

RULES ON CONFIDENTIALITY, SECURITY AND RELEASE OF INFORMATION

FOR USERS OF NHS PATIENT DATA

1) If the data received from NHS Tayside are to be held on computer, the signatory of this request, or the organisation (s)he represents, should have an appropriate registration with the Office of the Data Protection Registrar. Details of the registration number should be entered on this document.

2) Data received from NHS Tayside must not be used for any purpose other than for the intended use specified on this document.

3) Data received from NHS Tayside must not be divulged to any person who is not specified as a 'co-user of the data' on this document.

4) Proper safeguards should be applied in keeping the data and destroying it on completion of the work/project declared to prevent any breach of confidentiality.

5) Any misuse or loss of these data should be notified immediately to the Information Governance Officer for NHS Tayside at Ashludie Hospital, Monifieth (01382-527920).

6) Recipients of information supplied by NHS Tayside are reminded that the data has been supplied for the purposes stated in the approved study only. Further submission for approval will be required for any other uses of that data.

7) Any statistics or results of research based on data received from NHS Tayside should not be made available in a form which:
   a) directly identifies individual data subjects
   b) is not covered by the 'intended use of data' specified

NHS Tayside would welcome copies of any publications based on data supplied.

Information Governance
Ashludie Hospital
Monifieth
DD5 4HQ

Telephone : 01382 527920
Fax : 01382 527808

APPENDIX D: PUBLICATIONS AND PRESENTATIONS

All of the following publications/ presentations have been carried out during the course of my studies based on the original work carried out for this thesis.

Publications:

THE INFLUENCE OF FIELD STRENGTH AND DIFFERENT CLINICAL BREAST MRI PROTOCOLS ON THE OUTCOME OF TEXTURE ANALYSIS USING FOAM PHANTOMS
Shelley A. Waugh, Richard A. Lerski, Luc Bidaut, Alastair M. Thompson
Medical Physics (2011):38;5058-5066

Presentations:

COULD TEXTURE ANALYSIS BE A USEFUL TOOL IN THERAPEUTIC DECISIONS? PRELIMINARY RESULTS
SA Waugh, RA Lerski, AM Thompson
Due to be presented at the International Conference on Medical Physics, Brighton, 4th Sept 2013

REPRODUCIBILITY OF APPARENT DIFFUSION COEFFICIENT MEASURMENTS: RELIABILITY IN A CLINICAL SETTING
SA Waugh, Z Davis, S Vinnicombe
Due to be presented at the International Conference on Medical Physics, Brighton, 4th Sept 2013

EMERGING TECHNIQUES IN BREAST MRI
SA Waugh
Invited talk at UK Radiological Congress, Liverpool 10th June 2013

THE USE OF TEXTURE ANALYSIS IN BREAST MAGNETIC RESONANCE IMAGES
SA Waugh, RA Lerski, L Bidaut, AM Thompson
Presented at the College of Medicine, Dentistry and Nursing Annual Student Symposium; Dundee 5th June 2013

THE USE OF TEXTURE ANALYSIS IN IDENTIFICATION AND DIAGNOSIS OF BREAST CANCER ON MRI IMAGES
SA Waugh, RA Lerski, L Bidaut, AM Thompson
Presented at UK Radiological Congress, Manchester 25th June 2012

THE EFFECT OF CHANGING ACQUISITION CONDITIONS ON THE OUTCOME OF TEXTURE ANALYSIS USING CLINICAL BREAST MRI SEQUENCES ON A FOAM PHANTOM AT 1.5T AND 3.0T
SA Waugh, RA Lerski, L Bidaut, AM Thompson
Presented at the College of Medicine, Dentistry and Nursing Annual Student Symposium; Dundee 17th June 2011

Posters:

TEXTURE ANALYSIS OF MAGNETIC RESONANCE IMAGES IDENTIFIES EARLY CHANGES IN RESPONSE TO NEOADJUVANT CHEMOTHERAPY IN BREAST CANCER PATIENTS
SA Waugh, RA Lerski, AM Thompson
Presented at the National Cancer Research Institute, Liverpool, 3rd-6th Nov 2013

MAGNETIC RESONANCE IMAGING TEXTURE ANALYSIS DISTINGUISHED MOLECULAR SUBTYPES OF BREAST CANCER
SA Waugh, RA Lerski, AM Thompson
Presented at the National Cancer Research Institute, Liverpool, 3rd-6th Nov 2013

TECHNICAL COMPARISON OF TEXTURE ANALYSIS METHODS IN THE CLASSIFICATION OF BREAST CANCER SUBTYPES
SA Waugh, RA Lerski, AM Thompson
Presented at the International Conference on Medical Physics, Brighton, 4th-6th Sept 2013

THE USE OF TEXTURE ANALYSIS IN BREAST MAGNETIC RESONANCE IMAGING: DIFFERENTIATION OF DUCTAL AND LOBULAR CARCINOMA
SA Waugh, RA Lerski, L Bidaut, AM Thompson
Presented at the International Society of Magnetic Resonance in Medicine (ISMRM) conference; Melbourne 7th-11th May 2012

THE USE OF TEXTURE ANALYSIS IN THE GRADING OF BREAST CANCER ON MR IMAGES: PRELIMINARY FINDINGS
SA Waugh, RA Lerski, L Bidaut, AM Thompson
Presented at the International Society of Magnetic Resonance in Medicine (ISMRM) conference; Melbourne 7th-11th May 2012

THE USE OF TEXTURE ANALYSIS IN IDENTIFICATION AND DIAGNOSIS OF BREAST CANCER ON MRI IMAGES
SA Waugh, RA Lerski, L Bidaut, AM Thompson
Presented at the College of Medicine, Dentistry and Nursing Research Symposium; Dundee 26th Jan 2012

TEXTURE ANALYSIS IN IDENTIFICATION AND DIAGNOSIS OF BREAST CANCER USING MAGNETIC RESONANCE IMAGING

SA Waugh, RA Lerski, L Bidaut, AM Thompson

Presented at the College of Medicine, Dentistry and Nursing Annual Student Symposium; Dundee 15th June 2012

THE EFFECT OF ACQUISITION PARAMETER CHANGES ON THE OUTCOME OF TEXTURE ANALYSIS USING A CLINICAL BREAST MRI SEQUENCE ON A FOAM PHANTOM AT 1.5T

S Waugh, R Lerski, A Thompson

Presented at the International Society of Magnetic Resonance in Medicine (ISMRM) conference; Montreal 9th-13th May 2011

THE INFLUENCE OF FIELD STRENGTH AND DIFFERENT CLINICAL BREAST MRI PROTOCOLS ON THE OUTCOME OF TEXTURE ANALYSIS USING FOAM PHANTOMS

S Waugh, R Lerski, A Thompson

Presented at the International Society of Magnetic Resonance in Medicine (ISMRM) conference; Montreal 9th-13th May 2011

THE EFFECT OF CHANGING ACQUISITION CONDITIONS ON THE OUTCOME OF TEXTURE ANALYSIS USING CLINICAL BREAST MRI SEQUENCES ON A FOAM PHANTOM AT 1.5T AND 3.0T

SA Waugh, RA Lerski, L Bidaut, AM Thompson

Presented at the College of Medicine, Dentistry and Nursing Research Symposium; Dundee 27th Jan 2011